2013-06

An investigation of possible discriminating earned value variables in Department of Defense major acquisition program cancellation

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

AN INVESTIGATION OF POSSIBLE DISCRIMINATING EARNED VALUE VARIABLES IN DEPARTMENT OF DEFENSE MAJOR ACQUISITION PROGRAM CANCELLATION

by

Sidney W. Hodgson III

June 2013

Thesis Co-Advisor:	Diana Angelis
	Laura Armey
Second Reader:	Eugene Paulo

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE				*Form Approved OMB No. 0704–0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704–0188) Washington, DC 20503.				
1. AGENCY USE ONLY *(Leave blank)*		**2. REPORT DATE** June 2013		**3. REPORT TYPE AND DATES COVERED** Master's Thesis
4. TITLE AND SUBTITLE AN INVESTIGATION OF POSSIBLE DISCRIMINATING EARNED VALUE VARIABLES IN DEPARTMENT OF DEFENSE MAJOR ACQUISITION PROGRAM CANCELLATION				**5. FUNDING NUMBERS**
6. AUTHOR(S) Sidney W. Hodgson III				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943–5000				**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A				**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. government. IRB Protocol number ____N/A____.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited				**12b. DISTRIBUTION CODE** A
13. ABSTRACT (maximum 200 words)				
14. SUBJECT TERMS Earned Value Management, Major Defense Acquisition Programs, Cancellation, Mann Whitney, Probit, Probabilistic Modeling				**15. NUMBER OF PAGES** 133
				16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified		**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

Some major defense acquisition programs (MDAPs) are cancelled while others are not. The research evaluates earned value (EV) data reported by 12 cancelled MDAPs and compares it to the EV data reported by 12 MDAPs that had significant cost overruns yet were not cancelled (referred to as troubled programs). The study investigates whether there are differences in the key EV metrics reported for cancelled and troubled but not cancelled programs and seeks to develop a model that captures the probability of cancellation based on EV information. The thesis hypothesizes that cancelled programs would have more unfavorable cumulative cost and schedule variances, greater cost growth in the estimate at completion (EAC), and more disparity between the contractor and program manager cost estimates. Mann-Whitney tests were used to identify significant differences in EV metrics between the two groups and a probit model was used to further explore the relationship between EV information and the probability of program cancellation. The Mann Whitney and probit results, viewed together, suggest that early cost growth in a program may be an indicator of program cancellation. The thesis also suggests ways to improve EAC using probabilistic methods, including quantifying the level of cost risk associated with the EAC.

NSN 7540–01–280–5500 Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

AN INVESTIGATION OF POSSIBLE DISCRIMINATING EARNED VALUE VARIABLES IN DEPARTMENT OF DEFENSE MAJOR ACQUISITION PROGRAM CANCELLATION

Sidney W. Hodgson III
Commander, United States Navy
B.S., Villanova University, 1997

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN SYSTEMS ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
June 2013

Author: Sidney W. Hodgson

Approved by: Diana Angelis
Thesis Co-Advisor

Laura Armey
Thesis Co-Advisor

Eugene Paulo
Second Reader

Clifford Whitcomb
Chair, Department of Systems Engineering

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Some major defense acquisition programs (MDAPs) are cancelled while others are not. The research evaluates earned value (EV) data reported by 12 cancelled MDAPs and compares it to the EV data reported by 12 MDAPs that had significant cost overruns yet were not cancelled (referred to as troubled programs). The study investigates whether there are differences in the key EV metrics reported for cancelled programs and troubled but not cancelled programs and seeks to develop a model that captures the probability of cancellation based on EV information. The thesis hypothesizes that cancelled programs would have more unfavorable cumulative cost and schedule variances, greater cost growth in the estimate at completion (EAC), and more disparity between the contractor and program manager cost estimates. Mann-Whitney tests were used to identify significant differences in EV metrics between the two groups and a probit model was used to further explore the relationship between EV information and the probability of program cancellation. The Mann Whitney and probit results, viewed together, suggest that early cost growth in a program may be an indicator of program cancellation. The thesis also suggests ways to improve EAC using probabilistic methods, including quantifying the level of cost risk associated with the EAC.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACS	Aerial Common Sensor
ACWP	Actual Cost of Work Performed
ARH	Armed Reconnaissance Helicopter
ASDS	Advanced SEAL Delivery System
BAC	Budget at Completion
BCWP	Budgeted Cost of Work Performed
BCWS	Budgeted Cost of Work Scheduled
CAE	Component Acquisition Executive
CAIV	Cost as an Independent Variable
CDF	Cumulative Distribution Function
CEAC	Comprehensive Estimate at Completion
CPI	Cost Performance Index
CPR	Contract Performance Report
C/SCSC	Cost/Schedule Control Systems Criteria
CV	Cost Variance
CV %	Cost Variance Percentage
DAE	Defense Acquisition Executive
DAES	Defense Acquisition Executive Summary
DAMIR	Defense Acquisition Management Information Retrieval
DCMA	Defense Contract Management Agency
DFAR	Defense Federal Acquisition Regulation
DoD	Department of Defense
EAC	Estimate at Completion
EFV	Expeditionary Fighting Vehicle
EMD	Engineering Manufacturing and Development
ETC	Estimate to Complete
EV	Earned Value
EVM	Earned Value Management
EVMIG	Earned Value Management Implementation Guide
EVMS	Earned Value Management System

FAR	Federal Acquisition Regulation
FASA	Federal Acquisition Streamlining Act of 1994 Title V
FCS	Future Combat System
GAO	Government Accountability Office
GPRA	Government Performance and Results Act of 1993
GPS	Global Positioning System
IMS	Integrated Master Schedule
INCOSE	International Council on Systems Engineering
IPMR	Integrated Program Management Report
JHSV	Joint High Speed Vessel
JPALS	Joint Precision Approach and Landing System
JPATS	Joint Preliminary Aircraft Training System
MDAP	Major Defense Acquisition Program
NPOESS	National Polar Orbit Operating Environment Satellite System
OMB	Officer of Management and Budget
PAC	PATRIOT Advanced Capability
PARCA	Performance Assessments and Root Cause Analyses
PDF	Probability Density Function
PEO	Program Executive Office
PM	Program Manager
PM EAC	Program Manager’s Estimate at Completion
PMB	Performance Measurement Baseline
PMO	Program Management Office
RDT&E	Research, Development, Test and Evaluation
SAR	Selected Acquisition Report
SCI	Schedule Cost Index
SDD	Systems Development and Demonstration
SE	Systems Engineering or Systems Engineer
SPI	Schedule Performance Index
SV	Schedule Variance
SV %	Schedule Variance Percentage
TSAT	Transformational Satellite Communications

USD (AT&L)	Under Secretary of Defense for Acquisition, Technology and Logistics
VAC	Variance at Completion
WBS	Work Breakdown Structure
WSARA	Weapons System Acquisition Reform Act

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

PURPOSE

Some major defense acquisition programs (MDAPs) are cancelled while others are not. This thesis did not challenge the rationale behind MDAP cancellation; rather it sought to support the Under Secretary of Defense for Acquisition Technology and Logistics' (USD (AT&L)) directive to improve the effectiveness of earned value management (EVM) by providing additional insight into the relationship between unfavorable cost and schedule EVM variables and program cancellation. The primary objectives of this thesis were two-fold: (1) to investigate whether there were differences in the key earned value metrics of cancelled and troubled non-cancelled programs, including the effect that resetting variances may have on program survival, and (2) to develop a model that captures the probability of cancellation based on earned value information. The thesis also suggested ways to improve earned value cost estimation and reporting using probabilistic methods, including quantifying the level of cost risk associated with EV cost estimates. The primary benefit of this research is to provide Systems Engineers, Program Managers, and Department of Defense executives with additional insight into the relationship between unfavorable earned value metrics and program cancellation to improve decision making.

METHODOLOGY

This study looked at earned value data reported by 12 MDAPs that were eventually cancelled and compared it to the earned value data reported by 12 major acquisition programs that had significant cost overruns yet were not cancelled (referred to as troubled programs). The research investigated whether there were differences in the key earned value metrics reported for cancelled programs and troubled but not cancelled programs and sought to develop a model that captured the probability of cancellation based on earned value information. The thesis hypothesized that cancelled programs would have more unfavorable cost and schedule variances, greater cost growth in the estimate at completion (EAC), and more disparity between the contractor and program manager cost estimates. Mann-Whitney tests were used to identify significant differences in earned value metrics between the two groups, and a probit model was used to further explore the relationship between earned value information and the probability of program cancellation.

RESULTS

The data analyzed indicates that of the troubled programs sampled, cancelled programs do not have lower or more unfavorable cost or schedule variance percentages than non-cancelled programs. However, the Mann Whitney test results show that of the sampled programs, the cancelled programs have higher total cost growth over the life of the contract based on contractor EAC and program manager EAC. Probit modeling confirms that the significant variables are cumulative cost growth at 25% completion for both program manager and contractor EAC. Viewed together, the results suggest that there is reasonably strong evidence that early cost growth in a program is an indicator of program cancellation. Furthermore, a probit model (Figure i) was developed that captures the probability of cancellation based on earned value cost growth information.

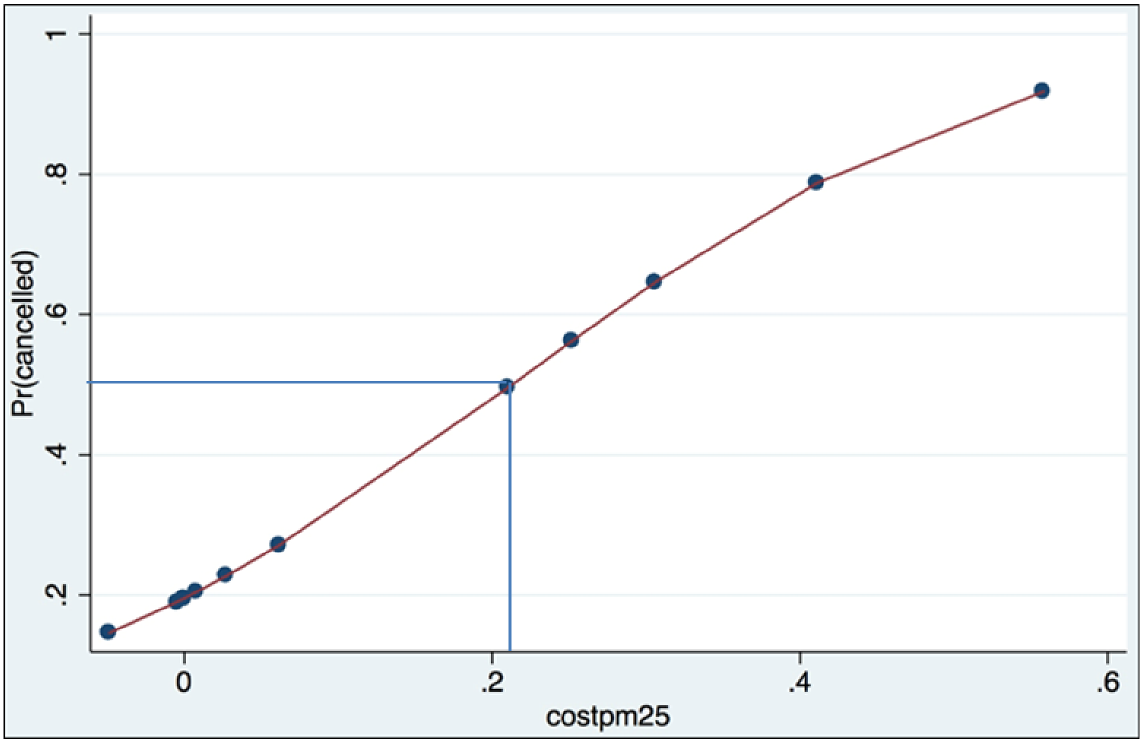

Figure i. Probit model using program manager cost growth at the 25% contract completion point.

Figure i depicts an S-curve that can be used to forecast the likelihood of program cancellation based on cost growth at the 25% contract completion point. For example, for cost growth greater than approximately 20%, the probability of cancellation is greater than 50%. This point is called the point of equal opportunity and is meaningful to program managers, systems engineers, and decision makers because it is the cost growth threshold where a program is more likely to be cancelled than not cancelled.

CONCLUSIONS AND RECOMMENDATIONS

The recommendations provided by this research are, (1) focused on the strongest finding of this report that cost growth in terms of EAC is an indicator of program cancellation, and (2) predicated upon the notion that probabilistic estimates that contain cost overrun risk information offer higher quality information and are more useful to decision makers (U.S. Government Accountability Office 2009). The three recommendations are:

1. Include more uncertainty information in the annual Comprehensive Estimate at Completion in the SAR reporting of MDAPs to Congress. The benefit of this additional information is that Congress would have a range of probabilistically determined EACs with associated levels of overrun risk to make cost growth assessments within programs and comparisons between programs.
2. Consider implementing a probabilistic approach to calculating estimates using existing earned value data throughout the life of the contract. This approach offers systems engineers and program managers additional insight into cost trends and overrun risk by providing monthly estimate and risk of overrun information with little additional effort required to construct the model.
3. Consider using probit modeling to better understand the impacts of early cost overruns on the likelihood of program cancellation.

The DoD has established Performance Assessments and Root Cause Analyses (PARCA) under USD (AT&L) to improve the effectiveness of EVM systems and to ensure that reliable, accurate, and timely EV data is transformed into realistic information that is useful to decision makers (Under Secretary of Defense for Acquisition, Technology and Logistics 2011). The above recommendations can help PARCA in their endeavor to achieve these improvements. All three of these recommendations improve the quality and realism of EV information presented to systems engineers, program managers, DoD officials and Congress and afford better management, budgeting, and cancellation decision making from a cost growth perspective.

ACKNOWLEDGMENTS

I must first express my sincere appreciation to my advisor and teacher, Dr. Diana Angelis. Her sound guidance, keen intelligence, and boundless energy were essential in helping me to navigate my way through uncharted territory. I'd also like to thank my other thesis committee members. Dr. Laura Armey's superb direction and novel approaches to solving problems were helpful when I got mired. Dr. Eugene Paulo was a terrific sounding board for ideas and shed light on interesting ways to see the problem and possible solutions. Thank you all.

I would also like to thank my fellow student, LCDR Ty Biggs. Working on thesis topics in similar areas allowed us to brainstorm different approaches and we proved that two minds are certainly three times as effective.

Finally, I would like to thank my family. To my son, Sidney, thanks for taking me to the park to play tee-ball and to feed the fish. At two years old, you created the balance that made this all possible. To Sofia, you came along at the end, but seeing you smile was always the highlight of my day. To my wife, Bridget, words cannot express what your love and support have meant to me. Thank you for your forbearance throughout this entire process. I am reminded everyday that you are the best decision I have made in my life. Thanks for picking me too. I love you all.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OVERVIEW

Twelve major defense acquisition programs (MDAPs) have been terminated since 2001 prior to fielding any operational systems (Harrison 2011). Former Secretary of Defense Robert Gates, in his April 2009 Defense Budget Recommendation Statement, stated that "the DoD must consistently demonstrate the commitment and leadership to stop programs that significantly exceed their budget or which spend limited tax dollars to buy more capability than the nation needs" (Gates 2009). He explains that numerous factors went into his cancellation recommendations, specifically cost overruns, schedule delays, and reprioritization of the capabilities to meet national interests. Similarly, when DoD Secretary Richard Cheney cancelled the Navy's A-12 program in January of 1991 he complained that no one could tell him how much it was going to cost. The more troubling fact was that there were numerous cost estimates that differed significantly and program managers were choosing to report only the most optimistic ones (Beach 1990).

In these austere economic times of increasing competition over scarce resources (the defense budget is expected to decrease over 20 percent, or roughly a trillion dollars, over the next 10 years [Adams 2013]), program managers face increasing pressure to achieve technical performance goals within established cost and schedule objectives. Programs reporting cost overruns, schedule delays and increasing or uncertain final total costs will be subject to more intense scrutiny and could increase their risk of termination.

This thesis focuses on Earned Value Management (EVM) and its role in assisting Systems Engineers (SEs), Program Managers (PMs), and Department of Defense (DoD) executives in evaluating the performance of MDAPs.

B. BACKGROUND

Earned Value Management is a widely accepted industry best practice that is used commercially and in the DoD to manage programs. It is an integrated management approach that uses schedule, cost, and scope of work goals and measures progress towards achievement of these goals (Performance Assessments and Root Cause Analyses

2013). EVM is a DoD Systems Engineering (SE) technical management process and an essential tool for proactive decision making (Defense Acquisition University 2012b). It provides systems engineers and Program Management Offices (PMOs) with information to manage the technical development of a system and to measure progress against an established baseline. EVM is required for all MDAPs. MDAPs are programs designated by the Under Secretary of Defense (Acquisition, Technology, and Logistics) (USD (AT&L)) as having an estimated eventual total expenditure for Research, Development, Test and Evaluation (RDT&E) of more than $365 million in FY 2000 constant dollars or, for procurement, of more than $2.19 billion in FY 2000 constant dollars (Defense Acquisition University 2012b). DoD has deemed Earned Value (EV) information critical to a program's success. USD (AT&L), the chief weapons buyer for the DoD, recently stated that, "EVM is one of DoD's and industry's most powerful program management tools" (Under Secretary of Defense for Acquisition, Technology and Logistics 2011).

Decision makers use EV information to evaluate programs' performance and to make critical decisions that align shrinking defense budget dollars to national security priorities. EV information is used as a measure of program progress and performance by Congress, the Government Accountability Office (GAO), DoD executive level decision makers including Program Executive Offices (PEO), Component Acquisition Executives (CAE), Defense Acquisition Executives (DAE), and USD (AT&L). The Weapons Systems Acquisition Reform Act of 2009 (WSARA), Public law 111-23, acknowledged the need for improvements to EVM and established the Office of the Secretary of Defense (OSD) Performance Assessments and Root Cause Analyses (PARCA) to provide additional EVM oversight and reporting on MDAPs (U.S. Congress 2009). Interestingly, this law also mandates the use of confidence level reporting for MDAP baseline estimates and annual updates, but not for earned value cost estimates.

In a recent DoD memorandum entitled "Earned Value Management Systems Performance, Oversight, and Governance," USD (AT&L) provides direction to improve the effectiveness of EV across the DoD. Specifically, he reemphasizes that EVM must be applied in a disciplined manner and that the data provided by EVM must be accurate, reliable and timely (Under Secretary of Defense for Acquisition, Technology and Logistics 2011).

C. PURPOSE

This thesis does not challenge the rationale to cancel the 12 MDAPs; rather it seeks to support USD (AT&L's) directive to improve the effectiveness of EVM by providing additional insight and perspective into the relationship between unfavorable cost and schedule earned value variables and program cancellation.

1. Research Objectives

The primary objectives of this thesis are two-fold: (1) to investigate whether there are differences in the key earned value metrics of cancelled and troubled non-cancelled programs, including the effect that resetting variances may have on program survival, and (2) to develop a model that captures the probability of cancellation based on earned value information. Based on these objectives the thesis looks at the following research questions:

- Are there statistically significant differences between cost growth, PM and Contractor cost estimation discrepancies, cost variance, and schedule variance in cancelled and troubled non-cancelled programs?
- Can a model be developed that predicts the likelihood of cancellation based on available EV data?
- Does resetting cost and schedule variance affect whether a program is cancelled?

This thesis hypothesizes that cancelled programs would have more unfavorable cost and schedule variances, greater cost growth in the estimate, and more disparity in estimates between the contractor and program manager cost estimates. Additionally, this thesis hypothesizes that non-cancelled programs re-set their variances more often than their cancelled counterparts.

2. Scope and Limitations

Figure 1 presents the high level view of the complicated Defense Acquisition environment with its major stakeholders—the user, industry, the executive branch, Congress, and Program Manager—along with the product and services provided by the respective stakeholder (the direction of the arrow indicates the stakeholder providing the product or service).

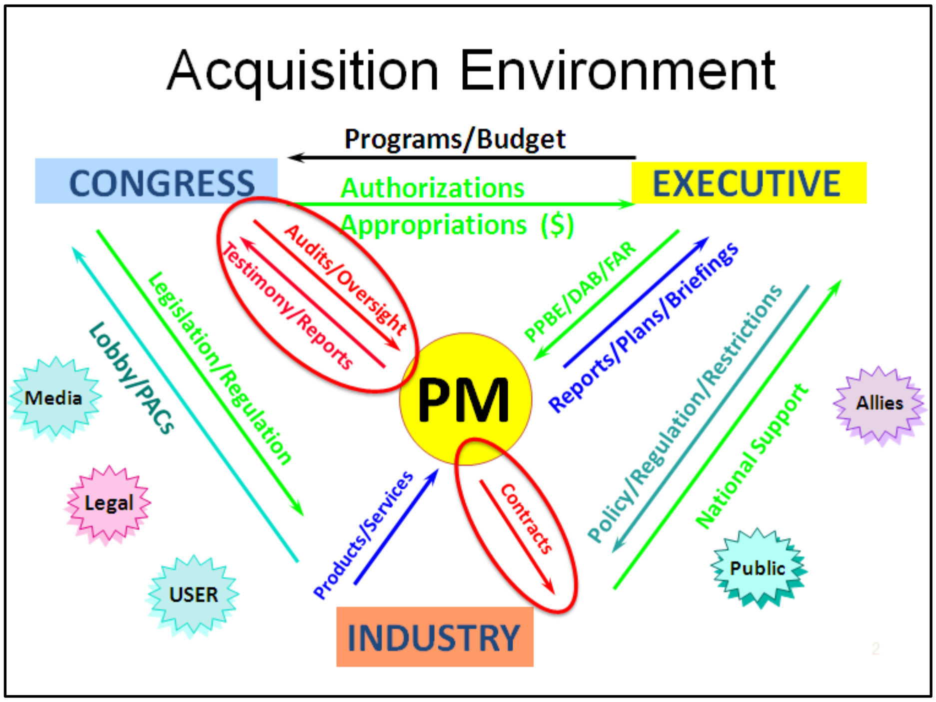

Figure 1. High-level view of defense acquisition environment (from Fast 2013).

The acquisition environment possesses many significant, diverse, and often competing interests, each of which could have a positive or negative effect on a program's success. This thesis focuses on investigating the earned value information that is supplied to Congress in their oversight role and DoD decision makers via official reports. The two domains circled in red in Figure 1 define the boundaries of this research. This thesis does not intend to determine causation; rather it investigates whether there is an association between unfavorable EV metrics and program cancellation.

Due to time constraints and restrictions on access to contractor data, this study is limited to the MDAPs cancelled since 2001 and programs that could be described as "troubled" but not cancelled in the same period of time. MDAPs were chosen because they are statutorily required to report EV data. This fact instilled confidence that sufficient data would be available for analysis. This thesis uses the EV data reported in the Defense Acquisition Executive Summaries (DAES) on the government's Defense Acquisition Management Information Retrieval (DAMIR) website. The DAES database contains the Contract Performance Report (CPR) data for all MDAPs.

This study focuses on unclassified MDAPs in the development phase of the Defense Acquisition Management framework. Figure 2 shows the evolution of the Defense Acquisition System from 1996 to the present and highlights the development phase (outlined in black). Note that the development phase is now called the Engineering & Manufacturing Development (EMD) phase, but from 2003 to 2008 it was referred to as the System Development & Demonstration (SDD) phase.

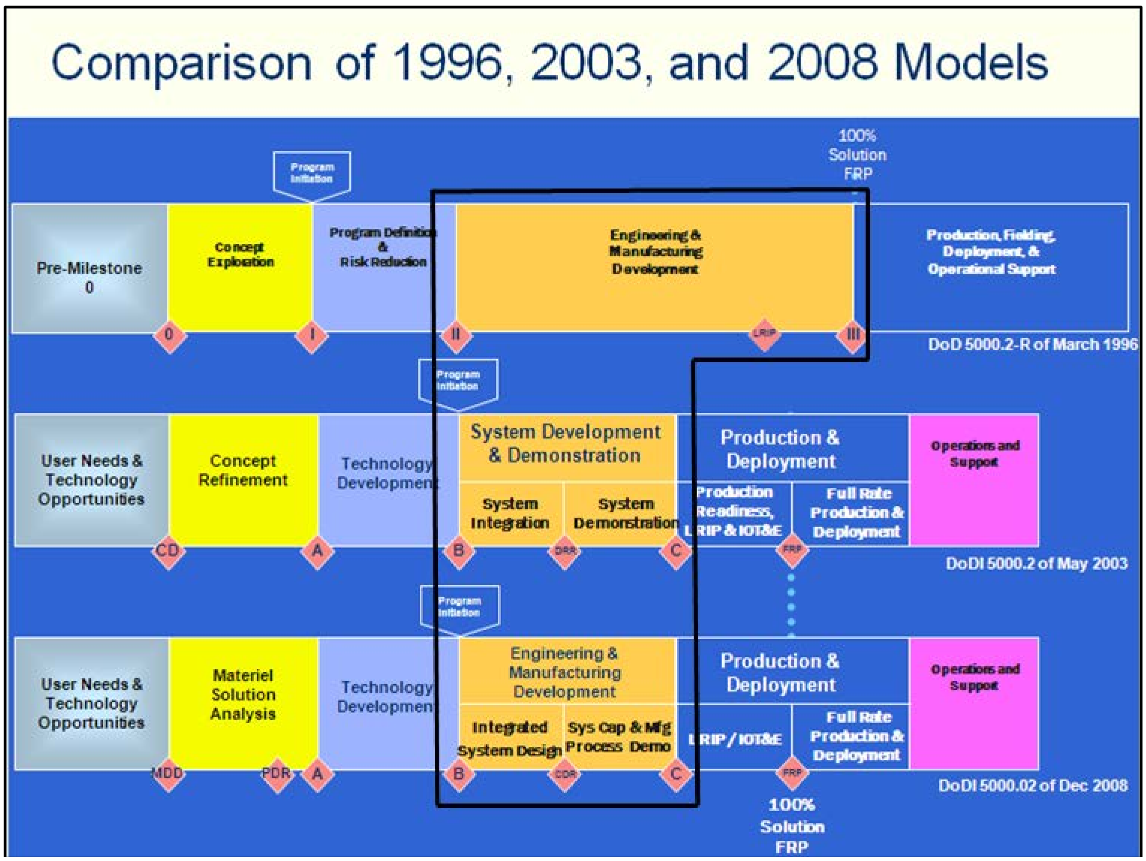

Figure 2. Comparison of Defense Acquisition Management Frameworks from 1993 to the present (from Fast 2013).

The comparison group of “troubled” programs was chosen using the following criteria. First, only contracts in the developmental phase (between Milestone B and C or between Milestone II and III of Figure 2) were considered. All of the cancelled programs were in this phase so it made sense to choose non-cancelled programs from this phase also. Second, only the largest EMD or SDD contract listed for each program in the DAES was used. All MDAPs have multiple contracts for each phase of weapons systems development to handle different components and functions of systems acquisition. To maintain consistency between the samples, the largest EMD/SDD phase contract was used for each program. Third, the program had to be “troubled.” Troubled programs were defined as those who have reported successive cumulative unfavorable cost or schedule variances worse than 10%. Amplifying information on program selection is presented in Chapter III.

3. Expected Benefit

EV information can be used to evaluate program performance and to make comparisons between multiple programs’ performance. The primary benefit of this research is to provide SEs, PMs, and DoD executives with additional insight into the relationship between unfavorable earned value metrics and program cancellation to improve decision making. Moreover, this research attempts to develop a model for forecasting the likelihood of cancellation based on existing EV metrics. This model will be helpful as a tool to help managers and decision makers better understand the possible effect unfavorable performance measures are having on the likelihood of program cancellation.

D. LITERATURE REVIEW

Numerous studies have been conducted that investigate why MDAPs fail. Most of the studies that were reviewed point to cost overruns as the primary cause and then evaluate the root causes of those overruns. The most cited reasons include: (1) excessively low initial cost estimates (U.S. Government Accountability Office 2009 and Sipple, White, and Greiner 2004), (2) expectation of cost overrun sharing causing firms to bid below cost estimate (Chen and Smith 2001), (3) excessively low cost overrun estimates (Christensen 1994), 4) technological immaturity (Dubos, Saleh, and Braun, 2007 and Tyson, Harmon, and Utech 1994), 5) unstable requirements (Augustine 1997), and 6) overly optimistic schedules (Berteau et al. 2011 and Augustine 1997). A 2008 GAO report on program cancellation found that of the 78 programs evaluated, only 11 of them were on schedule, on cost, and meeting technical performance requirements (Charette 2008).

Charette (2008) astutely points out that defense acquisitions problems have existed for decades; what has changed is the economic scope. To put the scope of defense spending in context, the DoD's 2012 portfolio of 86 MDAPs has a total estimated cost of $1.6 trillion (U.S. Government Accountability Office 2013); the DoD spends roughly $21 million an hour to acquire weapons systems. The cost overruns of the cancelled programs in this thesis alone total nearly $50 billion in "then year" costs while failing to field any of the systems. Clearly, cost and schedule overruns are an issue in the DoD. Berteau, et al. (2011) investigated the root cause of cost and schedule overruns and found that inaccurate cost estimates are associated with 40% of the accumulated cost overruns. In his thesis, Eric McKsymick (1995) found that the cost overruns in the cancelled A-12 program were excessive when compared to other programs. These findings sparked this thesis' fundamental research question: are there unfavorable earned value characteristics that distinguish cancelled programs from troubled non-cancelled programs? This is the fundamental question this research attempts to answer.

In addition to the numerous statutes, federal regulations, policies, and directives (additional detail for these is presented in Chapter II) requiring use of EVM in MDAPs, there is a great body of literature that advocates the use of EVM as a tool to help program managers, systems engineers, and decision makers monitor early signs of cost growth and take corrective actions as necessary (see for example Kerzner 2009 and Webb 2003). Abba (1997) makes it clear that the U.S. government and countries across the world have validated EVM as a highly effective program management tool.

There are numerous studies that discuss ways to improve EVM performance metrics. Langford and Franck (2009) acknowledge EVM as a useful tool to monitor defense acquisition and propose applying gap analysis and value analysis to make EV equations more realistic. The GAO (U.S. Government Accountability Office 2009) and Garvey (2000) both conclude that since costs are uncertain, estimating costs at completion as point estimates alone is insufficient for good decision making. They recommend quantitative and risk uncertainty analysis as a way to assess the variability in an estimate. This provides a level of certainty (and corresponding risk) with the estimate so that decision makers can better evaluate the estimate and informs them on cost, schedule, and technical risks (U.S. Government Accountability Office 2009).

Christensen has done extensive research analyzing EVM. He acknowledges that EV provides program managers and contractors valuable insight into the cost and schedule status of the project and has concluded through multiple studies that, (1) the estimate at completion (EAC) is one of the most critical values reported to PMs and that deficiencies in calculating and reporting EAC can adversely affect Congressional resourcing decisions and result in funding shortfalls (Christensen 1999), (2) of the numerous methods for determining EAC (both point estimates and regression models) some methods are more accurate at different completion points of the contract but no one method is always superior (Christensen 1993), and (3) methods for evaluating the reasonableness of the contractor's estimated cost at completion using EV data reported in monthly Contract Performance Reports should be used (Christensen 1999). While Christensen's analyses are very informative, these remain evaluations of deterministic point estimates. To reiterate, the GAO and Garvey have both highlighted the problem with basing decisions on point estimates alone because they do not provide a decision maker with any level of certainty information other than the estimate is the most likely (U.S. Government Accountability Office 2009 and Garvey 2000). They both advocate the use of cost uncertainty analysis as a way of introducing levels of certainty to estimates.

Ricardo Vargas (2009) suggests a cost uncertainty approach to determine EACs. He utilizes the three EAC point estimates (best case, worst case, and most likely case) for each Work Breakdown Structure (WBS) element in a triangular distribution, runs a Monte Carlo simulation and sums the element EAC results to estimate the programs' total EAC. This provides decision makers probabilistic EACs with associated levels of risk of overrun. This granularity of the EAC offers decision makers more insight into the final cost of the program and offers greater utility in program assessment during annual budget decisions. More details about this method are provided in Chapter II.

Extensive research and experience have convinced both Congress and USD (AT&L) to conclude that EVM is a highly effective program management tool (U.S. Congress 2009 and Under Secretary of Defense for Acquisition, Technology and Logistics 2011). Like all tools, however, it can be improved. The research above has given the DoD some ideas for improvements. DoD's recent commitment to improvements is codified in WSARA; WSARA created PARCA to conduct and oversee performance assessments and root cause analyses of MDAPs (U.S. Congress 2009). Moreover, the recent USD AT&L EVM governance memo reaffirms the DoD's commitment to EVM. USD AT&L has made PARCA responsible and accountable for EVM performance, oversight, and governance and for leading EVM improvements across the DoD (Under Secretary of Defense for Acquisition, Technology and Logistics 2011). Since its inception, PARCA has led the effort to implement EVM changes across the DoD. The most significant change involved improving the Integrated Program Management Report (IPMR). After assessing that cost and schedule information were oftentimes evaluated separately, the new IPMR integrates the cost information contained in the CPRs and schedule information in the Integrated Master Schedule (IMS) into one document to support consistent and robust program management (Office of the Assistant Secretary of Defense for Acquisition 2012). This thesis attempts to provide additional insight into other possible EVM improvements that may enhance the effectiveness of EVM in the DoD.

E. SUMMARY

This thesis examines the EV data from cancelled and non-cancelled MDAPs since 2001. This chapter presented the purpose and objectives of this research. Numerous factors are responsible for program cancellation. While many programs are troubled, not all of them are cancelled. This thesis explores whether there are EV characteristics that are more strongly associated with those cancelled programs. This study is organized into seven chapters. Chapter II provides additional background information on the importance of EV in Systems Engineering and program assessment, explains key EVM terms and performance metric calculations, and presents the merits of cost uncertainty methods. Chapter III presents details regarding data collection and Chapter IV explains the analytical approach. The results are presented in Chapter V. Chapter VI offers some ideas for using the results to improve program decision making, including a probabilistic modeling approach that provides cost risk information with the estimate. Chapter VII provides conclusions and recommendations for future research.

II. BACKGROUND

A. OVERVIEW

This thesis focuses on investigating Earned Value Management information available in statutorily required program reporting and its relationship to cancelled programs. Additionally, this thesis attempts to develop a model for DoD executives, program managers, and systems engineers that can help forecast the likelihood of cancellation based upon these same EV metrics. This chapter focuses on: (1) providing a context for the research including the importance of earned value in Systems Engineering and program management, (2) presenting a brief overview of the applicability of EV to the programs used in this study, and (3) defining the EV terms and performance metrics that will be used in the analysis.

B. THE ROLE OF EARNED VALUE MANAGEMENT IN SYSTEMS ENGINEERING AND PROGRAM MANAGEMENT

To understand the role of EV within Systems Engineering it is helpful to first define SE. Numerous definitions of Systems Engineering exist in the literature. The DoD has adopted the International Council on Systems Engineering (INCOSE) definition:

> Systems Engineering is an interdisciplinary approach and process encompassing the entire technical effort to evolve, verify and sustain an integrated and total life cycle balanced set of system, people, and process solutions that satisfy customer needs. Systems Engineering is the integrating mechanism for the technical and technical management efforts related to the concept analysis, materiel solution analysis, engineering and manufacturing development, production and deployment, operations and support, disposal of, and user training for systems and their life cycle processes. (Defense Acquisition University 2012b)

Systems Engineering uses numerous technical and management processes to evolve, integrate, sustain, and dispose of its systems. These processes are presented in Table 1.

Technical Management Processes	Technical Processes
Decision Analysis	Stakeholders Requirements Definition
Technical Planning	Requirements Analysis
Technical Assessment	Architectural Design
Requirements Management	Implementation
Risk Management	Integration
Configuration Management	Verification
Technical Data Management	Validation
Interface Management	Transition

Table 1. DoD Systems Engineering Technical Management and Technical Processes (adapted from Defense Acquisition Guidebook).

All of these processes in Table 1 are coordinated and executed at some point throughout the system's lifetime. It is beyond the scope of this thesis to describe each SE process and its value; however, it is necessary for this research to provide some detail about EVM's role and value within Systems Engineering technical management processes.

Systems engineers, especially those in management positions, program managers, and defense acquisition executives need to be aware of a project's status with respect to cost, schedule and performance. The Technical Assessment processes help provide these key decision makers with the information they need to assess, solve problems, and make important decisions regarding a program. Moreover, Technical Assessment activities measure technical progress within schedule and cost constraints. Earned Value Management is an important Technical Assessment process that uses earned value[1] systematically as the primary tool for integrating cost, schedule, and technical performance requirements (Kerzner 2009). It is a critical tool for engineering management and oversight of acquisition (Defense Acquisition University 2012b).

[1] Earned value is the value of completed work expressed as the value of the performance budget assigned to that work (Defense Acquisition University 2012b)

In simplest terms, EVM is a procedure for understanding, assessing, and quantifying what a contractor is achieving with contract dollars and to predict future performance. It works by establishing an integrated baseline that is developed from the work defined in the work breakdown structure and its associated time-phased budget. As work is performed, its corresponding budget ("earned value") can be measured against the integrated baseline. Cost and schedule variances can be calculated and analyzed. These variances can help management determine if a project is ahead or behind schedule and above or below budget, and where to focus additional resources to remedy the problem. The following is a list of a few of these EVM benefits:

- Provides insight into the contractor's performance
- Objectively measures work progress that is auditable
- Identifies significant cost and schedule drivers which can assist in determining the root cause of problems
- Forecasts future costs and scheduled completion date
- Facilitates communication

It is important to note that these variances reflect where problems occurred in the past. EVM can also use the current schedule and cost situation to forecast the project's end results based on its performance to date (Eisner 2002). Both aspects (reverse and forward looking) of EV variables are considered in this study.

The relevant statutes, regulatory policy, and DoD guidelines establishing earned value management as a key Technical Assessment process and valuable tool for decision makers are presented in the following section.

C. EARNED VALUE MANAGEMENT APPLICABILITY

1. EVM Statutes, Policy, and DoD Implementation

It is useful to this study to understand the relevance of EVM to Major Defense Acquisition Programs, its origin and regulatory policy, and its importance as an information tool for decision systems engineers, program managers, DoD executive, and Congress. EVM evolved during the 1990s from service unique cost and schedule performance criteria called the Cost/Schedule Control System Criteria (C/SCSC) into a set of 32 industry-owned guidelines called the Earned Value Management System (EVMS) (Defense Acquisition University 2012c). The requirement for MDAPs to use EVM is stipulated in three laws:

- Government Performance and Results Act of 1993 (GPRA)
- Federal Acquisition Streamlining Act of 1994 Title V (FASA)
- Clinger Cohen Act of 1996

These acts provide the legal basis for the policies implementing EVM. At the executive branch level, the primary policy document governing EVM is the Office of Management and Budget (OMB) Circular A-11 Part 7 Capital Programming Guide which requires the use of EVM or some similar system “for risk and program management of capital asset acquisition” and “to establish cost, schedule, and performance goals for major acquisitions and then achieve on average, 90% of these goals” (Office of Management and Budget 2006). The DoD, as an agency of the federal government, has issued its own directives (the DoDD 5000.01 (Department of Defense 2007) and DoDI 5000.02 (Department of Defense 2008)) to address the statutory and regulatory requirements of acquisition of all military systems. These documents along with the Federal Acquisitions Regulation (FARs) and the Defense Federal Acquisitions Regulation (DFARs) provide the explicit requirements[2] for use of EVM in DoD MDAPs.

The Weapon System Reform Act of 2009 established the Director, Performance Assessments and Root Cause Analyses (PARCA) and charged the organization with conducting and overseeing performance assessments and root cause analyses for MDAPs (U.S. Congress 2009). PARCA’s assessments evaluate the cost, schedule, and performance requirements relative to current metrics (Defense Acquisition University 2012b). Additionally, PARCA is the policy holder for EV; they are responsible for the implementation and use of EVM across the DoD and for evaluating the utility of performance metrics used to measure cost, schedule, and performance of MDAPs (Office of the Under Secretary of Defense for Acquisition, Technology and Logistics 2011).

[2] Current DoD regulation and policy requires EVMS on cost or incentive contracts, subcontracts, and intra- government work agreements valued at or greater than 20 million in then-year dollars. For efforts exceeding 50 million in then-year dollars, the EVMS must be validated or accepted by the Defense Contract Management Agency (Defense Acquisition University 2012b).

The DoD has recently acknowledged the need to improve the effectiveness of EVM in MDAPs. In his latest EVM Memorandum, the Under Secretary of Defense for Acquisitions, Technology, and Logistics (USD AT&L) offered that this new guidance will improve the effectiveness of EVM across the Department and to be successful "EVM practices and competencies must be integrated into the program manager's acquisition planning and execution processes; the data provided by EVM must be accurate, reliable, and timely and implemented in a disciplined manner" (Office of the Under Secretary of Defense for Acquisition, Technology and Logistics 2011). This memorandum continued the Defense Contract Management Agency's (DCMA) responsibility for EVM System compliance within the DoD. In this role, DCMA conducts EVMS reviews of all MDAPs to ensure compliance of EV standards. The DCMA's *Earned Value Management Implementation Guide* (EVMIG) is the DoD's principle guidance document for EV.

2. EVM Reporting Requirements

The three primary vehicles for reporting EVM information are: (1) Selected Acquisition Reports (SARs) to Congress[3], (2) Defense Acquisition Executive Summary (DAES)[4] reports to senior level DoD decision makers[5], and (3) the Integrated Program Management Report (IPMR)[6] to program managers. The goal of all of these reports is to facilitate communication between and provide feedback to key stakeholders in Congress,

---

[3] DoD must submit Selected Acquisition Reports (SARs) for all MDAPs annually. The frequency is increased should the MDAP fail to achieve certain performance thresholds contained in 10 USC § 2432 - Selected Acquisition Reports. The SARs enable USD (AT&L) to meet statutory reporting requirements of all MDAPs to Congress (Defense Acquisition University 2012b).

[4] DAES are submitted quarterly or monthly depending on whether certain performance thresholds are met. The DAES process enables the USD (AT&L) to fulfill statutory requirements to manage and oversee MDAPs. The goal of the DAES process is to facilitate communication between, and provide feedback to, key stakeholders in OSD, the Joint Staff, the Components, and Program Offices. It is important to note that the DAES is an internal management system meant to fulfill the needs of senior Department of Defense executives and is NOT for general public consumption (Defense Acquisition University 2012b).

5 For the remainder of this thesis the general collective term "DoD executives," "DoD officials," or "senior level DoD decision makers" refers to USD AT&L, Program Executive Office (PEO), the Milestone Decision Authority (MDA), and their associated staffs.

[6] The IPMR provides performance data that is used to identify problems early in the contract and forecast future contract performance (Defense Acquisition University 2012b).

DoD, and the program offices. Typically, the EV data for MDAPs is reported monthly by the contractor in Contract Performance Reports (formats 1 through 5)[7] and in the IPMR for use by internal program management. A portion of this data is included monthly in the DAES database, quarterly in the formal DAES reports for use by DoD executives, and annually in the SARs for use by Congress.[8] The DAES reports are the source documents for EV data in this research. Additional information pertaining to data collection and organization is presented in Chapter III.

D. EARNED VALUE TERMINOLOGY USED IN THIS RESEARCH

1. Earned Value Fundamentals

The following definitions are necessary to understand the analysis and findings of this report. The definitions come from the *EVMIG* (U.S. Department of Defense, Defense Contract Management Agency 2006) and the *Defense Acquisition University Glossary of Defense Acquisition and Terms* (Hagan 2009).[9]

- Budgeted Cost of Work Scheduled (BCWS or "planned value"): the sum of the budgets for all work scheduled to be accomplished with a given time period. Also called the Performance Measurement Baseline (PMB). $BCWS_{CUM}$ represents the cumulative BCWS at a certain point of the contract.
- Actual Cost of Work Performed (ACWP or "actual costs"): the costs actually incurred and recorded in accomplishing the work performed within a given time period. $ACWP_{CUM}$ represents the cumulative ACWP at a certain point of the contract.
- Budgeted Cost of Work Performed (BCWP or "earned value"): the value of completed work in terms of the work's assigned budget. $BCWP_{CUM}$ represents the cumulative BCWP at a certain point of the contract.
- Schedule Variance (SV): the algebraic difference between earned value and the budget (SV = BCWP – BCWS). A positive value is a favorable condition (ahead of schedule) while a negative value is unfavorable (behind schedule).

[7] Contract Performance Reports formats 1 through 5 are prepared by the contractor and are the primary means for reporting contract performance data. Their periodicity is typically monthly unless tailored for specific program.

[8] These periodicities are subject to change based on the program.

[9] Except where specific citation is made in this section it is to be assumed that the definition came from one of these two sources.

- Schedule Variance Percentage (SV%): indicates how much ahead or behind schedule the project is in terms of percentage. A positive value is a favorable condition (percent ahead of schedule) while a negative value is unfavorable (percent behind schedule). It may be expressed as a value for a specific period of time or for cumulative to date. SV % = (SV/BCWS) * 100 (TutorialsPoint 2013).

- Schedule Performance Index (SPI): EV performance factor representing schedule efficiency. Calculated by dividing the Budgeted Cost for Work Performed (BCWP) by the Budgeted Cost for Work Scheduled (BCWS). This metric is one of the performance factors used in EAC calculations.

- Cost Variance (CV): the algebraic difference between earned value and actual cost (CV = BCWP – ACWP). A positive value indicates a favorable condition (under budget) and a negative value indicates an unfavorable condition (over budget).

- Cost Variance Percentage (CV%): indicates how much over or under budget the project is in terms of percentage. It indicates how much less or more money has been used to complete the work as planned in terms of percentage. A positive value is a favorable condition (percent under budget) while a negative value is unfavorable (percent over budget). It may be expressed as a value for a specific period of time or for cumulative to date. CV % = (CV/BCWP) * 100 (TutorialsPoint 2013).

- Cost Performance Index (CPI): EV performance factor representing cost efficiency. Calculated by dividing the Budgeted Cost for Work Performed (BCWP) by the Actual Cost of Work Performed. This metric is one of the performance factors used in EAC calculations.

- Budget at Completion (BAC): The sum of all budgets established for the contract. BAC is a term that may also be applied to lower levels, such as the PMB or at the control account level.

- Estimate at Completion (EAC): the estimated total cost for all authorized work. Equal to the sum of actual costs to date (including all allocable indirect costs), plus the estimated costs to completion (estimate to complete).

- Estimate to Complete (ETC): estimate of costs to complete all work from a given point in time to the end of the contract.

- Variance reset: when a contract’s cost and/or schedule variances are reset to zero. This is done to improve managerial control over the work remaining on a contract.

When the BCWS, ACWP, and BCWP are obtained for a period, numerous additional EV metrics can be calculated including the SPI, SV%, CPI, CV%, ETC and EAC, that are helpful to understand how the program is performing and to predict future performance based on the contract’s past performance. Figure 3 illustrates the relationship of BCWS, ACWP, and BCWP for a project that is over-budget (ACWP>BCWP) and behind schedule (BCWP<BCWS).

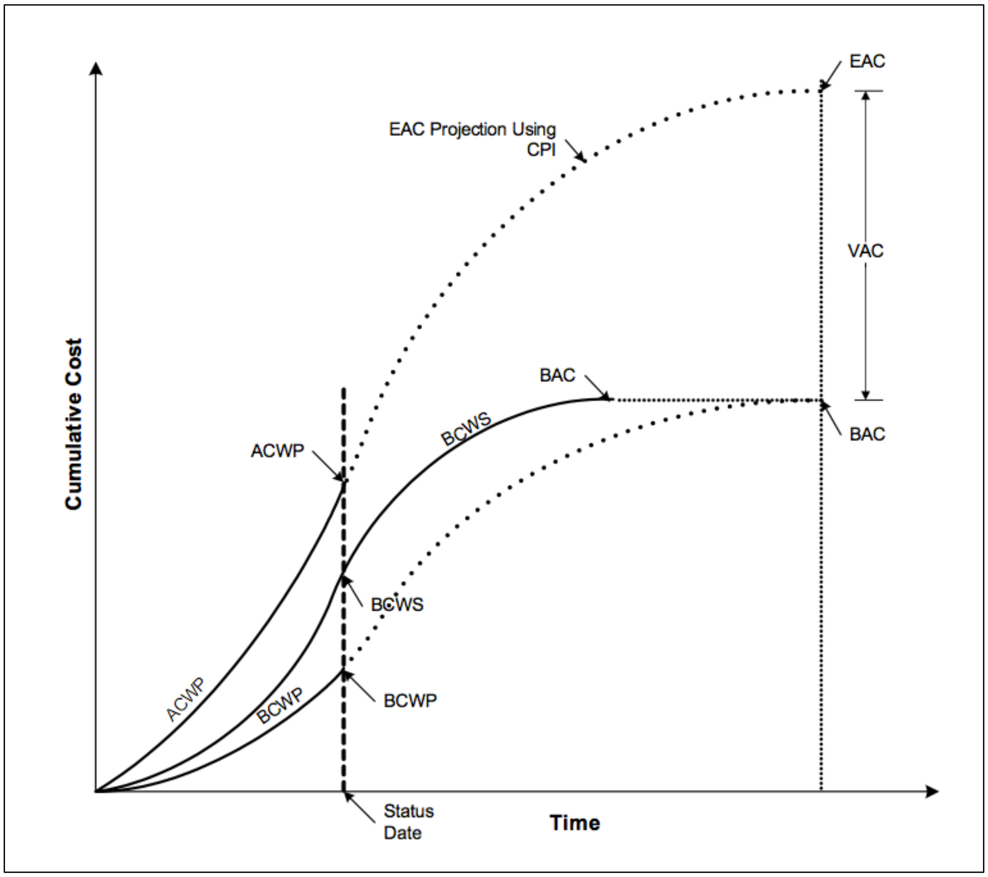

Figure 3. Graphical depiction of EV metrics for a sample project (from Vargas 2009).

In general, this research studies the relationship between CV%, SV%, EAC growth (as defined in Chapter III), and cost and schedule variance resets with program cancellation. All of these metrics are available in the DAES and considering the time limitation on this research their availability for all programs influenced their selection for this study. BAC was not available for every program, so analysis of Variance at Completion (VAC = BAC – EAC) for contracts was not possible. More detail about these performance metrics is contained in the chapters that follow. Since there are numerous ways to calculate EAC and it is not standardized in EV guidance documentation, the following section provides background on the two primary approaches the DoD uses to determine EAC.

2. Different Approaches to Calculating EAC

EAC is critical information for program managers and senior level decision makers. PMs use it to determine whether sufficient funds are available to cover the cost of the contract at completion. It also provides them insight into whether the costs are growing or not. USD (AT&L), Congress, and PARCA can also monitor the EAC growth rate and use this information in assessments and budgetary decision making. The DoD currently uses both deterministic and probabilistic methods for determining EAC for use in formal reporting. Both methods are explained further in the following sections.

a. Deterministic EAC Methods

A common way for programs to calculate their EAC is by using formulas that use contractor's efficiency to date as measured by the CPI and SPI. The resulting EACs are point estimates with no level of certainty associated with the estimate. Equation (1) illustrates that EAC is equal to the amount of money already spent on the contract or Actual Cost of Work Performed (ACWP) plus the amount of money it will take to complete the contract or the Estimate to Complete (ETC).

$$EAC = ACWP + ETC \tag{1}$$

The ACWP is an accounting figure reported monthly by the contractor in the CPRs. The ETC on the other hand is a forecast that can be calculated in numerous ways. The generic formula for calculating ETC is contained in Equation (2):

$$ETC = \frac{BAC - BCWP_{CUM}}{Index} \tag{2}$$

Use of different performance indices results in different EAC forecasts. The two indices used primarily in the DoD (but not mandated) are cost performance index and composite index. Use of these indices results in the EAC_{CPI}, referred to as the "best case" EAC, and the $EAC_{Composite}$, which is called the "worst case" EAC (Defense Acquisition University 2012a). These EAC values are reported as best and worst case EACs in the monthly DAES reports.

(1) Best Case EAC. The best case approach for calculating EAC involves using the current CPI of the program as the performance index. The reason it is best case is because it results in a lower EAC than its worst cast counterpart. It assumes that the rest of the work remaining will be done according to the same cost efficiency recorded to date. Equation (3) depicts this relationship:

$$ETC_{CPI} = \frac{BAC - BCWP_{CUM}}{CPI} \tag{3}$$

Substituting ETC_{CPI} into Equation (1) results in the EAC_{CPI} expression in Equation (4).

$$EAC_{CPI} = ACWP + \frac{BAC - BCWP_{CUM}}{CPI} \tag{4}$$

(2) Worst Case EAC. The worst case approach for calculating EAC involves using a composite performance index called the schedule cost index (SCI). Equation (5) shows the SCI expression.

$$SCI = SPI * CPI \tag{5}$$

It assumes that the rest of the future work will follow the cost efficiency determined by the cost performance index (CPI), as well as the schedule efficiency determined by the scheduled performance index, generating the SCI. Equation (6) depicts this relationship:

$$ETC = \frac{BAC - BCWP_{CUM}}{SCI} \tag{6}$$

This approach incorporates a tendency for programs to perform with CPIs and SPIs less than one (an indication of inefficiency). The product of two indices less than one has the compounding effect of raising the ETC more than using the CPI alone and consequently results in a higher EAC forecasts than the best case approach (see Equation (7)).

$$EAC = ACWP + \frac{BAC - BCWP_{CUM}}{SCI} \tag{7}$$

It is important to note that these methods, while providing the program manager and senior level decision makers with a range of estimates, remain deterministic point estimates to a probabilistic scenario. The DoD recognizes the pitfalls of point estimates in cost estimating and has leveraged probabilistic modeling in the creation of the annual Comprehensive Estimate at Completion (CEAC).

b. Comprehensive Estimate at Completion EAC

The CEAC is the current estimated total cost of authorized work on a contract. It is required to be completed annually by contractors in order to remain compliant with the FAR and DFAR for EVMS, and it is reviewed by DCMA. The CEAC is similar to the best and worst case EAC in that it equals actual costs at a certain point in time plus the estimated costs to complete the remaining work on the contract. The CEAC differs, however, because it is not a formula based macro level estimation technique. Rather, it requires rigorous probabilistic cost estimating methods similar to those used for estimating initial program costs. The objective of the CEAC is to serve as the most realistic and "most likely" estimate of program costs. The DCMA states that, "the most likely EAC has a probability of occurring of 50%, i.e., (P=0.50)[10]" (Makielski 2009). CEACs provide decision makers with a point estimate, its likelihood, and the cost risk associated with the estimate. Its benefit to program managers and decision makers is that it provides the level of certainty information along with the estimate. Users of this information can make an assessment of the risk of overrun based on the information provided with the point estimate. Moreover, CEAC aids program managers and decision makers in performing their responsibilities including making Cost as an Independent Variable (CAIV) decisions and determining if sufficient funds have been programmed. (Makielski 2009). The following paragraphs detail how the CEAC is developed and used.

Formulation of the CEAC is not deterministic like the macro formula based methods detailed in the previous section. Instead, it follows a cost uncertainty analysis process to improve the quality of the EAC. Essentially, this approach uses stochastic modeling of cost account information with underlying distributions in a Monte Carlo simulation to produce a point estimate EAC that includes its associated overrun risk. The GAO Cost Estimating Guide (U.S. Government Accountability Office 2009)

[10] The most likely estimate (known as the mode) does not necessarily correspond to a cumulative probability of 50% unless the probability density function is symmetrical.

contains very detailed information on developing a realistic Probability Density Function (PDF) and Cumulative Distribution Function (CDF) (otherwise known as an S-curve) to estimate costs. The four key steps that are applicable to formulating credible EACs are as follows:

1. Determine cost drivers and associated risks
2. Develop probability distributions to model uncertainty
3. Perform uncertainty analysis using Monte Carlo simulation
4. Identify the probability associated with the point estimate

It is not within the scope of this thesis to go through all of the detail of this process[11]; rather the intent here is to highlight how this process is used to develop the CEAC. The first step of this process is to conduct sensitivity analysis to help identify the cost drivers within the contract. Next, distributions are selected for each WBS element at the lowest level possible. Various distributions can be used depending on the nature of the costs for the particular cost account. These distributions could be based on analogous historical cost data or subject matter expertise. The cost distribution reflects the level of uncertainty in the estimate. For example, the triangular distribution is based upon three values: minimum, most likely, and maximum. If the spread or variability between these values is large, then there is greater uncertainty in this estimate.

A Monte Carlo simulation selects a value at random from each of the lower level WBS distributions then sums the random values to arrive at a single value for the total EAC. This is repeated thousands of times to produce the EAC Probability Density Function for the contract. Figure 4 is the result of a notional model that was created using 10 WBS cost accounts with varying triangular distributions.

[11] For additional information regarding cost uncertainty analysis and selecting realistic distributions see Garvey's *Probability Methods for Cost Uncertainty Analysis: A System's Engineering Perspective* (Garvey 2000) or the *GAO's Cost Estimating and Assessment Guide* (U.S. Government Accountability Office 2009).

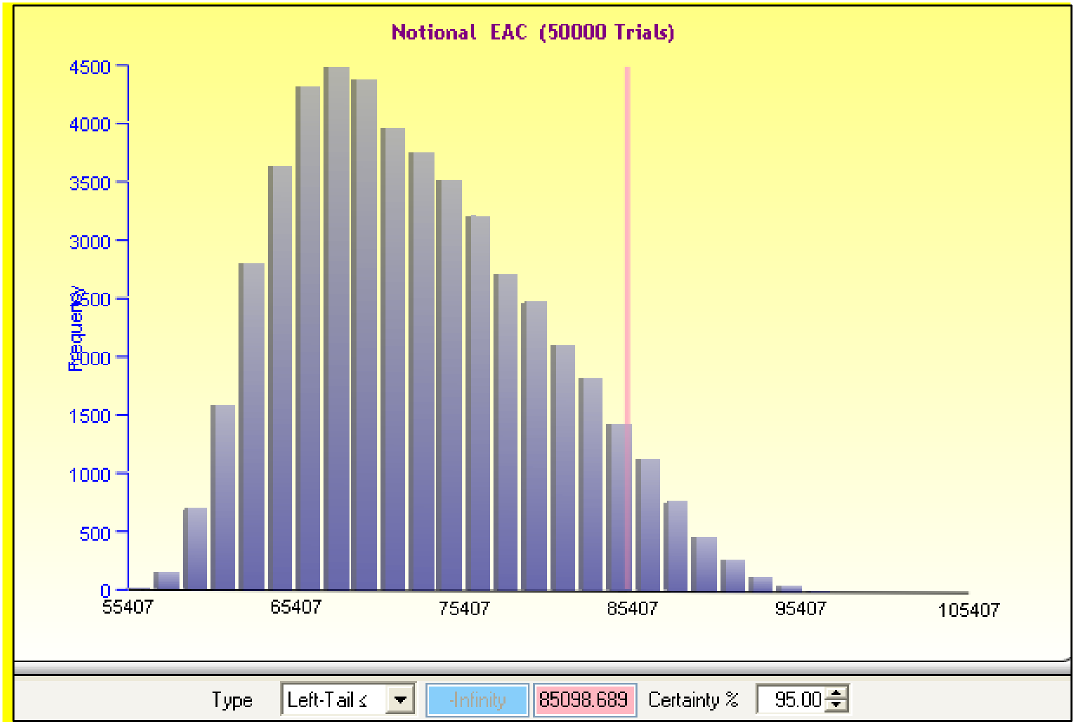

Figure 4. Notional CEAC (PDF) with 95% certainty percentage.

50,000 trials generated the PDF for the total contract seen in the figure. The one tailed 95% certainty threshold for this example is $85,099 and is indicated by a pink line on the PDF; this certainty level can be chosen based upon the manager's risk tolerance. Since the level of certainty is high (and risk of overrun is low), the resulting point estimate is higher and is similar to the worst case EAC calculated deterministically. Notice that decreasing the level of certainty results in a corresponding decrease in EAC and increase in cost overrun risk. An EAC towards the left edge of the curve would represent a "best case" EAC like the one mentioned above (the corresponding level of certainty for "best case" would need to be determined). The output of the model can also be displayed as a cumulative distribution function or S-curve. Figure 5 presents an example of a CDF with various cost estimates mapped to their corresponding probability of occurrence levels.

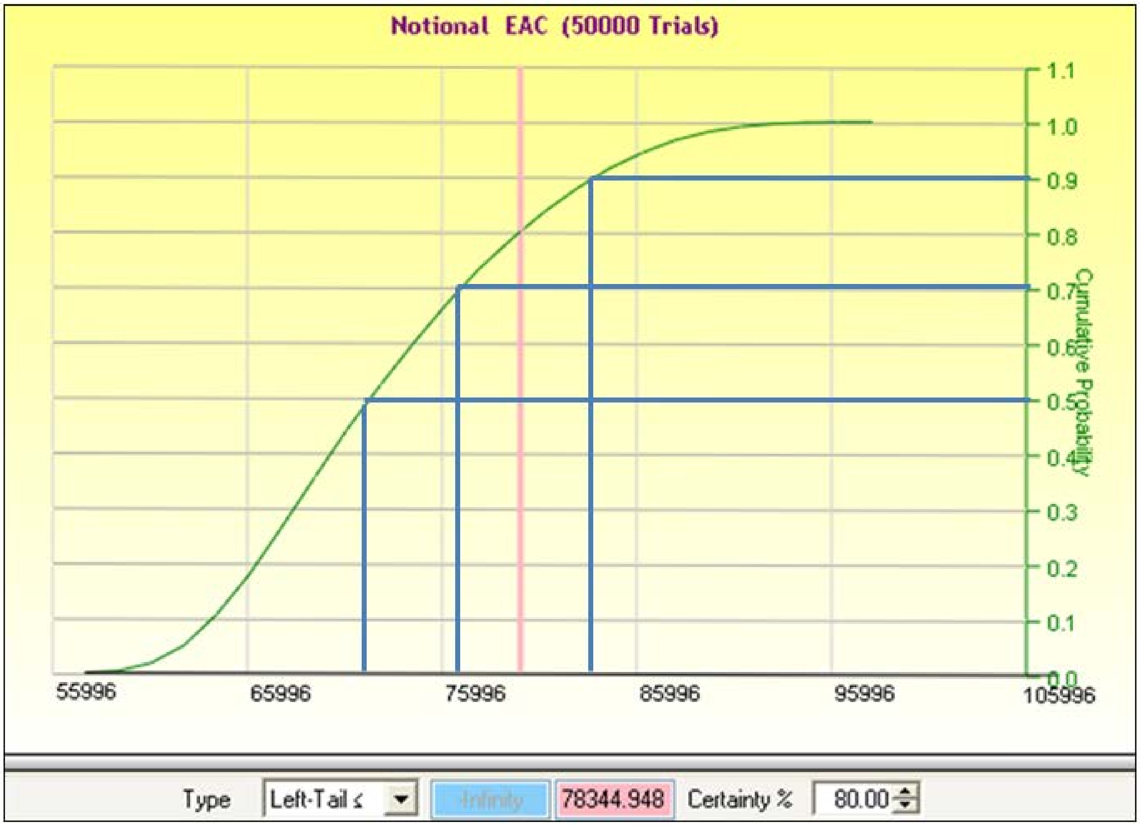

Figure 5. Notional CEAC S-curve (CDF).

At 50% certainty level, the estimated cost is approximately $70,000. As the certainty level is increased to 70%, the estimation increases to roughly $76,000, and to $78,344 at 80%. Finally, at 90% level of certainty, the estimate increases to roughly $81,000. This increase in the estimate can be thought of as the cost of certainty (Angelis 2012), i.e., how much more must be spent to reduce the risk of overrun.

CEAC is a more realistic estimate because it considers uncertainty in its creation instead of solely using past performance as the indicator of future costs. Moreover, it is useful in checking the deterministic estimates for realism. Monthly contractor EACs should fall within the bounds of these CEACs if the program is on track. If the contractor EACs do not fall within these bounds, then the root cause of the growth can be investigated. Finally, by comparing the recent CEAC to previous CEAC (at the same level of certainty), cost growth can be measured. This is a useful metric when making comparisons between programs.

Two values for EAC (Program Manager and Contractor) are reported in the SARs; neither the statute for SARs nor the SAR Data Entry Instructions state the associated level of certainty of the EAC used in SAR reporting. This thesis assumes that the “most likely” (probability = 0.50) CEAC is reported in the SARs.

E. SUMMARY

EVM is a system required by law and DoD policy and is useful to systems engineers, program managers, and senior level decision makers. It measures a contract’s cost and schedule performance, helps identify problem areas in a contract, and predicts future performance. WSARA recognized a need for improvements and established PARCA to provide additional oversight and reporting of MDAPs EV performance. As a result, the DoD has placed increased emphasis on improving EV measurement and analysis tools in MDAPs in its directives, regulations, and memorandums. This thesis seeks to determine if certain EV variables are associated with cancelled programs and to provide an additional tool that will aid in this EV analysis improvement effort.

THIS PAGE INTENTIONALLY LEFT BLANK

III. DATA

A. OVERVIEW

The purpose of this chapter is to familiarize the reader with the process used to obtain the data used in this study. The key objective of the data process was to find sufficient EV data to conduct the analysis and help answer the research questions. This chapter lists the programs used in the study and the criteria for selecting cancelled and non-cancelled programs. It then describes the process for collecting and organizing the EV data from the DAMIR database. Additionally, it provides basic descriptive statistics of the data collected and the rationale for the chosen analytical approach.

DAMIR is a DoD initiative that streamlines acquisition management and oversight by leveraging numerous government databases into one central repository. The DAMIR databases used in this study were the Selected Acquisition Report (SAR) and the Services' Defense Acquisition Executive Summary (DAES). The DAES information contains some of the earned value data from the contract performance reports (DD Forms 2734/1–5, APR 2005) on a nearly monthly basis (not all EV data available in the CPR formats 1–5 is available for study). The SARs, in the most extreme case, are published three times per year, but more typically are published once annually and do not contain sufficient EV data for meaningful analysis in this study; however, the SAR is useful because it provides background information on each program and some explanations of program actions taken. The analysis reported in this thesis is based almost exclusively on the DAES data because of the frequency and quality of the EV data provided in DAES.

B. PROGRAMS IN THE STUDY

As discussed previously in Chapter I, the primary objectives of this thesis are: (1) to investigate whether there are differences in the key earned value metrics of cancelled and non-cancelled programs, and (2) to develop a model that captures the probability of cancellation based on earned value information. The first task in achieving these objectives was to determine which programs, both cancelled and non-cancelled, to use in the study.

1. Cancelled Program Selection

Since 2001, 12 Major Defense Acquisition Programs were cancelled before they could field an operational system (Harrison 2011). These programs are listed in Table 2. Of the 12 cancelled programs, DAMIR earned value data was available for the eight programs annotated with an asterisk. These eight programs made up the cancelled programs sample of this study.

Program Name	Sunk Cost (in billions of then-year dollars)	Source
Future Combat Systems (FCS)*	$18.1 B	September 2008 SAR
Comanche*	$7.9 B	December 2003 SAR
National Polar-orbiting Operational Environmental Satellite System (NPOESS)*	$5.8B	December 2009 SAR
VH-71 Presidential Helicopter*	$3.7B	September 2008 SAR
Expeditionary Fighting Vehicle (EFV)*	$3.3B	December 2010 SAR
Transformational SATCOM (TSAT)	$3.2B	Annual Budget Justification Document
Crusader*	$2.2B	December 2001 SAR
Advanced SEAL Delivery System (ASDS)	$0.6B	December 2005 SAR
Armed Reconnaissance Helicopter (ARH)*	$0.5B	September 2008 SAR
Aerial Common Sensor (ACS)*	$0.4B	December 2005 SAR
CG(x) Next Generation Cruiser	$0.2B	Annual Budget Justification Document
CSAR-X	$0.2 B	Annual Budget Justification Document
* Program used in this study		

Table 2. Major programs cancelled since 2001 without fielding any operational systems (adapted from Harrison 2011).

2. Non-Cancelled Program Selection

This thesis aims to discern whether there are differentiating earned value characteristics within troubled programs and then use these characteristics to form a model to forecast the likelihood of program cancellation. Conceivably, more can be learned by comparing cancelled programs to fellow troubled programs than by comparison to on-track programs. One would expect the on-track programs to outperform the cancelled programs in most, if not all, earned value performance metrics. It is more informative to look for differentiators among troubled programs. Recall that the Performance Assessments and Root Cause Analyses (PARCA) in the Office of the

Assistant Secretary of Defense for Acquisition serves as the DoD's focal point for all policy, guidance and competency relating to EVM (Under Secretary of Defense for Acquisition, Technology and Logistics 2011). PARCA states that cumulative cost and schedule variance reportable thresholds should be +/- 5% (Office of the Assistant Secretary of Defense for Acquisition 2012). Additionally, recall from Chapter II that the OMB Circular requires that major acquisitions establish cost, schedule, and performance goals for major acquisitions and then achieve on average, 90% of them (Office of Management and Budget 2006). Also, it is a generally accepted DoD program management heuristic that unfavorable variances greater than 10% are considered significant and potentially troublesome. All three of these influenced this thesis' definition of troubled programs as those programs that experienced consecutive unfavorable (negative) cumulative cost or schedule variance percentages greater than 10%. The selection criterion used for the comparison sample of the non-cancelled programs was three-fold: (1) the program was not cancelled, (2) earned value data was available for the major program, and (3) the program had to be troubled (as defined above).

Due to time constraints and to maintain consistency in sample sizes, eight programs were selected from the 90 active programs in DAMIR for use as a control group. The programs are organized by service on DAMIR and then alphabetized within the service. The first row of each service was analyzed until eight programs that met the criteria were obtained. The first eight non-cancelled programs that met the criteria for troubled programs were chosen as the comparison sample of non-cancelled programs. These programs are listed Table 3.

Program Name
H1 Upgrades
Joint Precision Approach and Landing System (JPALS)
CH-53K
Global Positioning System (GPS) III
Joint High Speed Vessel (JHSV)
PATRIOT Advanced Capability-3 (PAC-3) System
UH-60M
Joint Primary Aircraft Training System (JPATS)

Table 3. Comparison sample of troubled non-cancelled major programs.

With the cancelled and non-cancelled samples determined, the next steps were to collect and organize the EV data for use in the analysis.

C. EARNED VALUE DATA: COLLECTION AND ORGANIZATION

The second task involved obtaining and organizing the EV data required to conduct the analysis and answer the research questions. This section reviews the variables studied, describes collection and organization of the data, and provides basic descriptive statistics.

This study hypothesized that the variables presented in Table 4 were the most likely differentiators of cancelled programs. In order to sufficiently analyze the hypotheses that these variables are different in cancelled programs, the following EV data was collected for each program (definitions for each variable are presented in Chapter II):

- Cost variance percentage
- Schedule variance percentage
- Program Manager’s Estimate at Completion (PM EAC)
- Contractor’s Estimate at Completion
- Cost variance resets
- Schedule variance resets

The collection, organization and descriptive statistics for each group of variables listed in Table 4 are discussed in the next section.

Independent Variables
Cost Variance %
Cost Variance % at 25% completion
Cost Variance % at 50% completion
Schedule Variance %
Schedule Variance % at 25% completion
Schedule Variance % at 50% completion
Cost Growth (PM Estimate at Completion)
Cost Growth (PM Estimate at Completion) at 25% completion
Cost Growth (PM Estimate at Completion) at 50% completion
Cost Growth (Contractor Estimate at Completion)
Cost Growth (Contractor Estimate at Completion) at 25% completion
Cost Growth (Contractor Estimate at Completion) at 50% completion
Difference between Contractor and Program Manager's EAC
Difference between Contractor and Program Manager's EAC (0-25%)
Difference between Contractor and Program Manager's EAC (26-50%)
Difference between Contractor and Program Manager's EAC (51-75%)
Cost Variance Reset Frequency
Schedule Variance Reset Frequency

Table 4. Variables used in analysis.

1. Contract Selection

All MDAPs have multiple contracts for each phase of the weapons systems development to handle different components and functions of systems acquisition. To maintain consistency between the samples, the largest Engineering and Manufacturing Development phase contract was used for each program. The available raw data for the variables for each program's largest EMD contract was extracted from the DAMIR site under the "Earned Value" tab using the "Cumulative" and "Summary" reports and stored into a comprehensive master data spreadsheet. While the DAES reports are missing some data, sufficient data existed for each program to conduct the analysis.

The raw data was then organized for analysis. There were different amounts of data available for each program and at various stages of contract completion. To handle this, all available data was collected for sample comparison and then the data was partitioned to test for timing of variances and growth. If the collected data was a percentage or a frequency then direct comparisons could be made between the samples. For the data that was not a percentage, the procedure for normalizing it is described in the following sections. The following paragraphs explain how the data was organized and discuss the descriptive statistics. The analytical methodology using two sample hypothesis testing and probit regression modeling is detailed in Chapter IV and the results of this analysis are presented in Chapter V.

2. Cost Variance Percentage

The cost variance percentages were extracted from every CPR available in the DAES reports. The mean and medians were calculated and are presented in Table 5. Interestingly, the overall median and mean CV% of non-cancelled programs is more unfavorable than the cancelled programs. This suggests that the non-cancelled programs are as troubled (and perhaps more so) than their cancelled counterpart. Additionally, it suggests that CV% may not be a discriminating variable in a program's survival. The differences were explored further using two-sample hypothesis testing, and the results are presented in Chapter V.

Cancelled program	Median CV%	Mean CV%	Std. Dev.	Non-cancelled program	Median CV%	Mean CV%	Std. Dev.
VH-71	0.12	-1.31	4.75	H1 Upgrades	-0.97	-2.66	7.16
FCS	0.58	0.63	1.38	JPALS	-6.28	-6.99	2.54
Comanche	-4.59	-4.31	3.45	CH-53K	-0.97	-1.50	4.02
NPOESS	-3.23	-4.58	4.90	GPS III	-5.62	-7.77	4.82
EFV	-6.93	-7.14	3.44	JHSV	-24.37	-24.64	13.09
Crusader	-1.94	-3.03	3.31	PAC-3	-5.54	-8.20	3.58
Armed Recon Helo	-4.58	-8.73	9.58	UH-60M	-5.47	-4.26	4.03
ACS	-1.94	-5.47	11.05	JPATS	-9.54	-8.13	12.50
Overall	-2.59	-4.24		Overall	-5.58	-8.02	

Table 5. CV percentage means and medians of all programs in study.

The same CV% data was utilized to analyze the potential impact of the timing of cost variance percentages on program cancellation at the 25% and 50% completion points of the contract. For these two completion points, the average of the CV% within +/- 5% of the respective completion points was used for the CV% of each program. Tables 6 and 7 contain the CV% at these completion points for the sampled cancelled and non-cancelled programs (note: data for VH-71 and ACS was not available in DAES for these contract completion points).

Cancelled program	CV% at 25% Contract Completion Point	CV% at 50% Contract Completion Point
VH-71		-13.01
FCS	1.12	-1.07
Comanche	-9.07	-4.96
NPOESS	1.11	-1.34
EFV	-13.19	-4.31
Crusader	-5.47	-6.16
Armed Recon Helo	-6.18	-21.37
ACS		
Median	-5.82	-4.96
Mean	-5.28	-7.46

Table 6. Cancelled programs' CV percentages at 25% and 50% contract completion points.

Non-cancelled program	CV% at 25% Contract Completion Point	CV% at 50% Contract Completion Point
H1 Upgrades	-11.13	-9.78
JPALS	-4.30	-6.19
CH-53K	-1.23	-5.16
GPS III	-4.55	-12.49
JHSV	-7.78	-10.96
PAC-3	-6.33	-8.15
UH-60M	7.84	-1.51
JPATS	14.85	-11.29
Median	-4.42	-8.96
Mean	-1.58	-8.19

Table 7. Non-cancelled programs' CV percentages at 25% and 50% contract completion points.

Table 6 suggests that on average cancelled programs have more unfavorable CV% at the later stages of the contract. However, the median CV% appears to improve from 25% to 50% completion point. Table 7 shows that both the mean and median CV%

becomes more unfavorable over time for the sampled non-cancelled programs. When comparing the mean CV% between cancelled and non-cancelled programs in Tables 6 and 7, cancelled programs maintain worse CV% at the 25% completion point, but slightly more favorable at the 50% completion point. These observations are not intuitive and make drawing any initial conclusions challenging.

3. Schedule Variance Percentage

The schedule variance percentages were extracted and compiled from every available CPR in the DAES reports. The mean and median for each program were calculated and are presented in Table 8. The mean SV percentage is slightly more unfavorable (more negative) in cancelled programs while the cancelled programs' median SV% is slightly less unfavorable.

Cancelled program	Median SV%	Mean SV%	Std. Dev.	Non-cancelled program	Median SV%	Mean SV%	Std. Dev.
VH-71	-1.94	-2.20	1.13	H1 Upgrades	-1.24	-2.67	2.74
FCS	-0.91	-0.94	0.45	JPALS	-2.31	-2.31	1.14
Comanche	-2.19	-2.10	1.75	CH-53K	-2.73	-3.16	2.59
NPOESS	-1.01	-4.44	6.79	GPS III	-3.41	-3.76	2.47
EFV	-3.71	-3.82	1.74	JHSV	-6.07	-6.05	2.47
Crusader	-0.98	-3.24	4.72	PAC-3	-1.85	-3.33	4.28
Armed Recon Helo	-19.78	-16.30	11.89	UH-60M	-7.86	-8.70	4.35
ACS	-5.86	-9.93	10.44	JPATS	-6.23	-6.19	5.69
Overall	-2.06	-5.37		Overall	-3.07	-4.52	

Table 8. SV percentage means and medians of all programs in study.

To analyze the timing of schedule variances and their potential effect on cancellation, the same SV% data was used. For the 25% and 50% completion points of the contract, the average of the SV% within +/- 5% of these completion points was used for the SV% of each program. Tables 9 and 10 contain the SV% at these completion points for the sampled cancelled and non-cancelled program.

Cancelled program	SV% at 25% Contract Completion Point	SV% at 50% Contract Completion Point
VH-71		
FCS	-0.93	-0.93
Comanche	-4.20	
NPOESS	-5.56	-0.36
EFV	-6.80	-4.44
Crusader	-10.51	-11.39
Armed Recon Helo	-30.98	-20.71
ACS		
Median	-6.18	-4.44
Mean	-9.83	-7.56

Table 9. Cancelled programs' SV percentages at 25% and 50% contract completion points.

Non-cancelled program	SV% at 25% Contract Completion Point	SV% at 50% Contract Completion Point
H1 Upgrades	-3.68	-3.08
JPALS	-3.31	-3.41
CH-53K	-1.52	-0.55
GPS III	-1.72	-6.03
JHSV	-8.99	-5.49
PAC-3	-16.13	-5.32
UH-60M	-25.63	-7.22
JPATS	-6.10	-6.77
Median	-4.89	-5.41
Mean	-8.38	-4.73

Table 10. Non-cancelled programs' SV percentages at 25% and 50% contract completion points.

Table 9 suggests that on average cancelled programs have more favorable SV% at the later stages of the contract. Table 10 shows that while the mean SV% seems to improve over time, the median SV% becomes more unfavorable for the sampled non-cancelled programs. When comparing the mean and median SV% between cancelled and non-cancelled programs in Tables 9 and 10, cancelled programs maintain worse SV% at the 25% completion point. There is conflicting evidence between the mean and median

SV% at the 50% point. Again, these general observations based on very small sample sizes make drawing initial conclusions challenging.

4. Variance Percentage Rate of Change

Figure 6 plots CV% and SV% with the percent complete of the contract for the Expeditionary Fighting Vehicle (EFV) program. This is an attempt to compare the rates of change of the cost and schedule variance for all sampled programs.

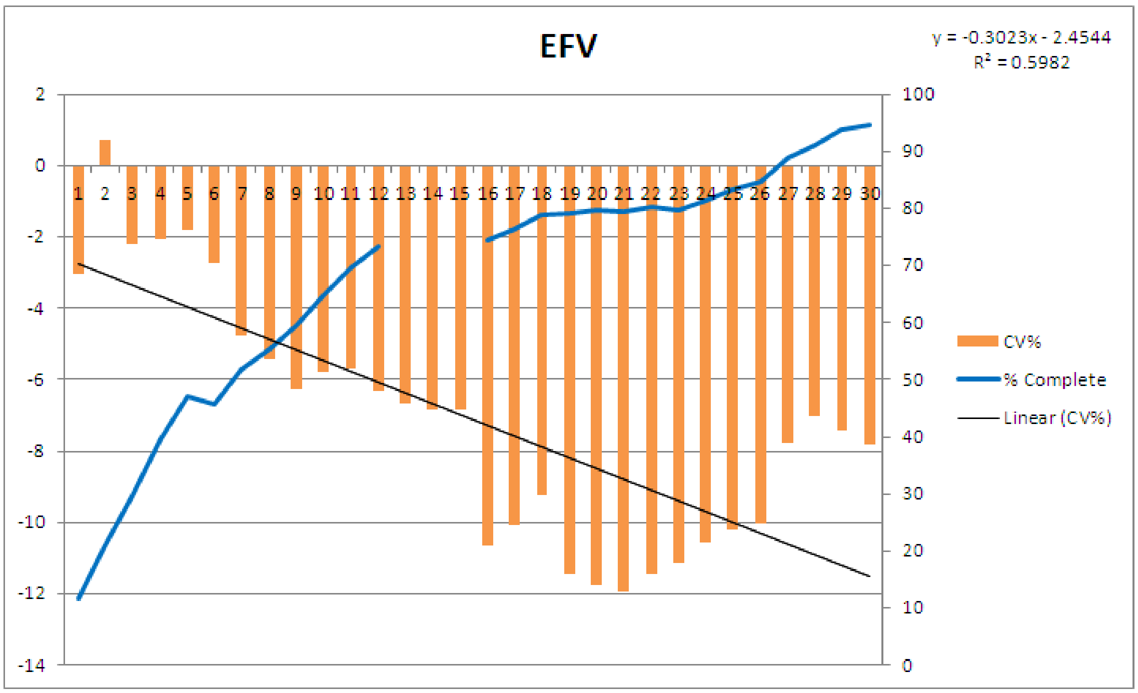

Figure 6. CV% rate of change for EFV program.

The left vertical axis represents the CV%; the right vertical axis is contract completion percentage. The contract's completion percentage is depicted by the blue curve (the non-continuous portion of the blue curve represents missing DAES contract completion percentage data). A trend line was fitted to the data and the corresponding slope of the line represents the estimated slope of the fitted line. The plan was to use these rates as an additional variable for analysis. However, most coefficients of determination or R^2 values for the cost and schedule variance percentage plots were below 0.5 and therefore inconclusive. Future research could explore these rates of change as potential variables for analysis.

5. Cost Growth

To analyze cost growth as a potential discriminating variable, program manager estimate at completion and contractor EACs were recorded for every CPR. Initially the cumulative cost growth, that is, the difference between the EAC at the beginning of the contract and the current EAC would seem the logical metric. However, due to changes in the program baseline ("re-baselining") during the course of the contract, the cumulative cost growth did not prove a consistent measure from program to program. When a program experiences significant cost growth or schedule delays, it is not uncommon for the program manager to request that it be "re-baselined." This gives the program a fresh start, and leads to a new "beginning" EAC. One possible way to handle the effect of re-baselining is to consider the re-baselined program a new program. This method was discarded because it was difficult to determine in some cases when or if a program had been re-baselined (and in some cases the EAC would go down without any evidence of re-baselining).

Alternatively, it is possible to ignore the re-baselined EAC and continue to calculate the difference between the original EAC and the current EAC. Doing so is problematic however, as it fails to recognize that some programs are re-baselined due to changes in scope or requirements, so that the observed cost growth is not indicative of program health, but merely a reflection of legitimate changes in program scope. Since the objective of this research was to find variables that may predict program cancellation, measuring the difference between the original EAC and the new EAC could be misleading.

To address the difficulties mentioned above, a cost growth metric was constructed to facilitate comparison of programs. The purpose of the metric is to measure the cumulative effect of marginal changes in the EAC. This metric has the advantage that it is less distorted by re-baselining yet still captures the overall effect of increases in the EAC from the beginning of the program to the current period. The metric for both PM EAC and contractor EAC was calculated using Equation (8):

$$\text{EAC Growth }\%_{present} = \text{EAC Growth }\%_{past} + \frac{\text{EAC}_{present} - \text{EAC}_{past}}{\text{EAC}_{past}} \text{ x } 100 \qquad (8)$$

Note that this constructed metric will always be lower than the cumulative cost growth percentage (which is computed by taking the difference between the current EAC and the original EAC and dividing the difference by the original EAC). This constructed metric provides a conservative estimate of the cost growth experienced by a program, while eliminating the problems caused by re-baselining. This metric is hereafter referred to as cost growth. For the 25% and 50% contract completion points, the average cost growths within +/- 5% of 25% and 50% completion points, respectively, were computed and used for the cost growth of each program. The total cost growths and cost growths at 25% and 50% completion points were calculated for all programs in the study and are presented in Tables 11 thru 14. Where data is missing it is because it was unavailable in DAES.

Cancelled program	Total Cost Growth (PMEAC)	at 25% Contract Completion Point	at 50% Contract Completion Point
VH-71	-36.64%		
FCS	55.28%	41.11%	47.23%
Comanche	93.48%	55.81%	
NPOESS	113.52%	25.13%	88.25%
EFV	86.31%	6.16%	23.15%
Crusader	399.29%	337.04%	258.43%
Armed Recon Helo	151.21%	-0.51%	23.92%
ACS	22.31%		
Median	89.90%	33.12%	47.23%
Mean	110.60%	77.46%	88.19%

Table 11. Cost growth for cancelled programs (PM EAC).

Non-cancelled program	Total Cost Growth (PMEAC)	at 25% Contract Completion Point	at 50% Contract Completion Point
H1 Upgrades	109.49%	20.99%	45.93%
JPALS	20.00%	0.72%	2.86%
CH-53K	33.03%	0.00%	0.01%
GPS III	-2.01%	0.00%	-0.04%
JHSV	27.17%	2.74%	2.74%
PAC-3	67.63%	-4.88%	30.01%
UH-60M	83.12%	0.00%	70.81%
JPATS	58.51%	30.62%	66.56%
Median	45.77%	0.36%	16.43%
Mean	49.62%	6.27%	27.36%

Table 12. Cost growth for non-cancelled programs (PM EAC).

Cancelled program	Total Cost Growth (Contractor EAC)	at 25% Contract Completion Point	at 50% Contract Completion Point
VH-71	20.04%		
FCS	55.30%	41.28%	50.04%
Comanche	98.81%	55.96%	
NPOESS	118.13%	24.02%	92.87%
EFV	84.54%	3.08%	27.86%
Crusader	393.44%	329.59%	338.79%
Armed Recon Helo	143.95%	-0.51%	52.83%
ACS	21.60%		
Median	91.68%	32.65%	52.83%
Mean	116.98%	75.57%	112.48%

Table 13. Cost growth for cancelled programs (Contractor EAC).

Non-cancelled program	Total Cost Growth (Contractor EAC)	at 25% Contract Completion Point	at 50% Contract Completion Point
H1 Upgrades	106.23%	19.49%	43.27%
JPALS	24.28%	2.45%	9.74%
CH-53K	37.65%	21.83%	32.14%
GPS III	0.60%	3.92%	17.82%
JHSV	27.05%	0.00%	2.06%
PAC-3	66.71%	-9.20%	22.86%
UH-60M	5.00%	0.00%	0.00%
JPATS	56.99%	5.13%	102.77%
Median	32.35%	3.19%	20.34%
Mean	40.56%	5.45%	28.83%

Table 14. Cost growth for non-cancelled programs (Contractor EAC).

Figure 7 provides a general illustration of the trend of cost growth of the sampled programs over time.

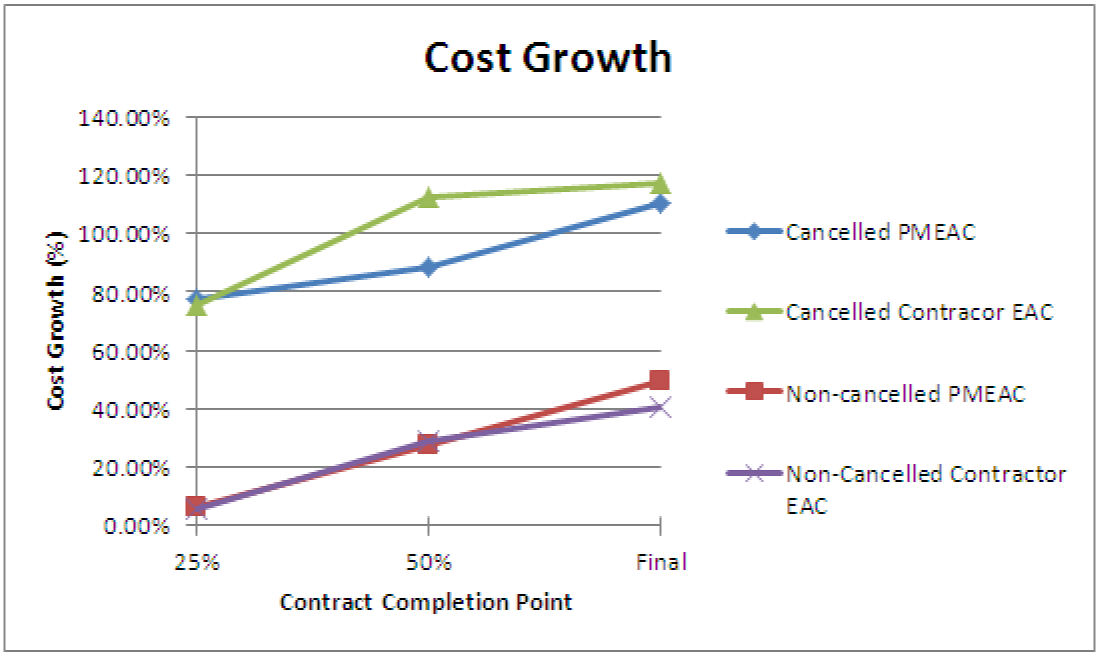

Figure 7. Cost growth of cancelled and non-cancelled programs over time.

Some general observations can be made from Tables 11 thru 14 and Figure 7. Expectedly, regardless of estimate type, as programs progress, their cost growth in the form of estimates at completion increases. Also note that cost growth appears to be significantly higher in cancelled programs than in the non-cancelled counterparts. Moreover, the data appears to show that the there are fewer differences in program manager and contractor estimates in non-cancelled programs and that the greatest difference in estimates appears to occur at the 50% completion point of cancelled programs.

6. Differences in the Estimates at Completion

To investigate the potential effect the differences between the two estimates may have on cancellation, it was first necessary to normalize the differences. The calculation to normalize the differences is based on the PM EAC and Contractor EAC values reported in each program's CPR in the DAES report. In order to normalize the differences for program comparison, the magnitude of the difference was divided by its corresponding program manager's estimate at completion for each program. For example, if the program manager's estimate at completion was $5000 and the contractor's estimate at completion was $4500 at a certain contract completion percentage, the difference between the two estimates is $500 and the normalized difference is $500/$5000 or 0.10.

The data was organized into the following four periods to study the potential effects of the timing of the difference: (1) total normalized difference, (2) normalized difference from 0–25% contract completion point, (3) normalized difference from 26–50% contract completion point, and (4) normalized difference from 51–75% contract completion point. The median total EAC difference and averages of the three periods' EACs were calculated for each program in the study and are presented in Tables 15 and 16. The cells of the table that remain empty did not have sufficient data for analysis.

Cancelled program	Total EAC Difference*	EAC Difference (from 0-25%)	EAC Difference (from 26-50%)	EAC Difference (from 51-76%)
VH-71	38.31%			32.16%
FCS	0.00%	0.00%	0.00%	0.00%
Comanche	3.77%	4.99%	5.87%	
NPOESS	3.07%	18.69%	12.01%	3.13%
EFV SDD1	1.39%	1.16%	2.51%	2.04%
Crusader	1.28%		1.19%	1.49%
Armed Recon Helo	0.00%	0.00%	0.61%	2.68%
ACS	2.80%	5.19%		
Median	2.09%	3.08%	1.85%	2.36%
Mean	6.33%	5.01%	3.70%	6.92%
* Median EAC Difference				

Table 15. Normalized differences in program manager and contractor EAC for cancelled programs.

Non-cancelled program	Total EAC Difference*	EAC Difference (from 0-25%)	EAC Difference (from 26-50%)	EAC Difference (from 51-76%)
H1 Upgrades	0.77%	0.15%	1.52%	5.34%
JPALS	2.34%	3.00%	4.49%	3.11%
CH-53K	7.24%	6.46%	9.36%	6.31%
GPS III	8.82%	9.67%	9.55%	9.11%
JHSV	0.13%	0.00%	4.51%	1.02%
PAC-3	0.87%	0.00%	1.23%	4.63%
UH-60M	4.67%	0.00%	1.16%	6.34%
JPATS	5.37%	8.37%	11.56%	2.31%
Median	3.51%	1.57%	4.50%	4.98%
Mean	3.78%	3.45%	5.42%	4.77%
* Median EAC Difference				

Table 16. Normalized differences in program manager and contractor EAC for non-cancelled programs.

The medians and means tell two different stories in this case. A comparison of the medians reveals that the EAC difference is worse (greater) in non-cancelled programs for all but the 0–25% group. Whereas, a comparison of the means suggests that cancelled programs have worse EAC differences in all groups but the 26–50%. Conclusions cannot be made from simply comparing the means and medians in this way, but it is useful in providing a general sense of how the difference in estimates may effect cancellation.

7. Cost and Schedule Variance Resets

Another variable considered as potentially discriminating was the frequency of cost and schedule variance resets. There was not an efficient way of doing this other than to manually count the resets in the EV data. At times the reset could be cross referenced to the DAES report or SAR for additional details, however, most of the time this was not possible. Table 17 and 18 present the number of variance resets for each program in the study. The low frequency of resets makes drawing conclusions challenging.

Cancelled program	CV Reset Frequency	SV Reset Frequency
VH-71	1	0
FCS	0	0
Comanche	1	0
NPOESS	2	2
EFV	0	0
Crusader	1	1
Armed Recon Helo	1	1
ACS	0	0
Median	1	0.00
Mean	1	0.50

Table 17. Cost and schedule variance reset frequency for cancelled programs.

Non-cancelled program	CV Reset Frequency	SV Reset Frequency
H1 Upgrades	2	2
JPALS	0	0
CH-53K	1	2
GPS III	0	0
JHSV	0	0
PAC-3	1	1
UH-60M	0	0
JPATS	0	0
Median	0	0.00
Mean	1	0.63

Table 18. Cost and schedule variance reset frequency for non-cancelled programs.

D. SUMMARY

This chapter explained the data collection in preparation for answering the research questions. It was difficult to draw any definitive conclusions from cursory analyses of the means and medians; nevertheless, it was helpful to observe some of the trends within the samples and possible relationships that may exist between the samples. It was critical at this point in the research process to determine the most appropriate analytical tools to help answer the research questions considering the small sample sizes and varying number of available data points per program. Chapter IV offers the details of the analytical approach and the rationale behind its selection. The results of the application of these analytical tools are then presented in Chapter V.

IV. ANALYTICAL APPROACH

The two stated objectives of this thesis are: (1) to investigate whether there are differences in the key earned value metrics of cancelled and non-cancelled programs, and (2) to develop a model that captures the probability of cancellation based on earned value information. This chapter develops the hypotheses and explains the selection rationale of the statistical techniques and their application in the analysis.

A. CANCELLED AND NON-CANCELLED PROGRAM COMPARISON

1. Hypotheses

This thesis examines whether there are differences in the earned value metrics of cancelled and non-cancelled programs that may be associated with program cancellation. For earned value metrics where a more negative value equates to more unfavorable performance, such as cost variance percentage and schedule variance percentage, the null and alternative hypotheses in expressions (9) and (10) applies:

$$H_0 : \mu_{\mathrm{C}} = \mu_{\mathrm{NC}} \tag{9}$$

$$H_A : \mu_{\mathrm{C}} < \mu_{\mathrm{NC}} \tag{10}$$

The null hypothesis in expression (9) states that the mean percentages for cancelled and non-cancelled programs are equal. The alternative hypothesis in expression (10) states that the mean percentage for the cancelled programs is less (more negative) than the mean percentage for the non-cancelled programs.

For earned value metrics where a higher value is more unfavorable, such as estimate at completion cost growth, difference between contractor and program manager EAC, and variance reset frequency, the null and alternative hypotheses are provided in expressions (11) and (12):

$$H_0 : \mu_{\mathrm{C}} = \mu_{\mathrm{NC}} \tag{11}$$

$$H_A : \mu_{\mathrm{C}} > \mu_{\mathrm{NC}} \tag{12}$$

The null hypothesis in expression (11) states that the mean for cancelled and non-cancelled programs is equal. The alternative hypothesis in expression (12) states that the mean for the cancelled programs is greater than the mean for the non-cancelled programs.

2. Mann Whitney Testing

The first objective of this thesis is to determine if earned value metrics exist that could help indicate cancellation. To this end, it was necessary to conduct hypothesis testing of the selected earned value variables. Data was available for only eight cancelled programs since 2001. There are several potential tests to analyze the available data, but the small sample sizes and the fact that the underlying distributions are most likely not normal, makes Mann Whitney, explained below, the best test for this task.

The first test considered was the two independent samples t-test. Typically, statistical t-testing would be used to test hypotheses such as those contained in this thesis. However, the standard t-test and other parametric hypothesis testing procedures are based on an assumption that the data is a random sampling from a normally distributed population. It is rather unwise to make this assumption of normality with such small sample sizes; in these cases, non-parametric hypothesis tests are more appropriate.

The Mann-Whitney test is the non-parametric equivalent to the independent samples t-test and can be used when you do not assume that the dependent variable is a normally distributed interval variable (University of California, Los Angeles: Statistical Consulting Group 2013b). The Mann Whitney test can be used to test for differences between the medians of two populations[12]. It was chosen since the data samples for cancelled programs (n = 8) and non-cancelled programs (n = 8) were very small and not assumed to be normally distributed. One-sided Mann Whitney tests were conducted with an alpha = 0.10 to determine if there are differences between the medians of the earned value variables in cancelled and non-cancelled programs. At this significance level, the chance of making a Type I error, i.e., rejecting the null when it is in fact true, is 10% or less. In other words, there is a 10% chance that the test would show a difference in the variable being studied in cancelled and non-cancelled programs, when in reality no difference existed. While increasing alpha may slightly increase the likelihood of incorrectly rejecting the null hypothesis, it does help detect whether a difference is

[12] Note that Mann Whitney uses medians, not means. It is a non-parametric statistical hypothesis test for assessing whether one of the two samples of independent observations tends to have larger values than the other.

present. All variables in Table 19 were tested using Minitab® software (Minitab is a statistical software package developed and distributed by Minitab, Inc.).

Independent Variables
Cost Variance %
Schedule Variance %
Cost Growth (PM Estimate at Completion)
Cost Growth (Contractor Estimate at Completion)
Difference between Contractor and Program Manager's EAC
Cost Variance % at 25% completion
Cost Variance % at 50% completion
Schedule Variance % at 25% completion
Schedule Variance % at 50% completion
Cost Growth (PM Estimate at Completion) at 25% completion
Cost Growth (PM Estimate at Completion) at 50% completion
Cost Growth (Contractor Estimate at Completion) at 25% completion
Cost Growth (Contractor Estimate at Completion) at 50% completion
Difference between Contractor and Program Manager's EAC (0-25%)
Difference between Contractor and Program Manager's EAC (26-50%)
Difference between Contractor and Program Manager's EAC (51-75%)
Cost Variance Reset Frequency
Schedule Variance Reset Frequency

Table 19. Variables tested using Mann Whitney.

The Mann Whitney tests produce p-values. The p-values for this test represent the probability of incorrectly identifying a difference between cancelled and non-cancelled programs. For example, when testing the cost variance percentage, if the resulting p-value is 0.05, the null hypothesis is rejected, but there is a 5% chance that the claimed difference is erroneous. For these tests, results with p-values less than the stated alpha = 0.10 are considered statistically significant, the null hypothesis is rejected, and the conclusion that there is a difference in the medians of the respective variable for cancelled programs could be drawn. This information could be used to assess whether certain earned value metrics are worse in cancelled programs than non-cancelled programs. This information is valuable to program managers, systems engineers and decision makers because it may confirm or refute long standing heuristics regarding earned value metrics and cancellation. It may also be valuable as an indicator of the variables that are most likely to be used in developing a probability regression model. More information on probit modeling is contained in the next section.

B. PROBIT REGRESSION MODELING

One of the objectives of this thesis is to determine whether earned value metrics can be used to establish a model that indicates the probability of program cancellation based on earned value data of the cancelled and non-cancelled programs. The results of the Mann Whitney analysis suggest drivers of differences that may be helpful in developing such a model. The model had to meet two main criteria to achieve the desired objective. Due the nature of the data, the model must be equipped to handle the binary response variable in this thesis, i.e., cancelled and not cancelled. The model must also be able to explain probability of cancellation.

While for many types of problems, linear regression models produce efficient estimators, the following disadvantages of linear regression models made them ill-suited for this thesis' purposes (Koop 2012):

- Linear regression models imply that the dependent variable is normally distributed; in the case of this thesis it is either 0 or 1 and not normal.
- The fitted value for the dependent value won't be 0 or 1 even though these are the only two values of the dependent variable
- Linear regression models return unreasonable estimated probabilities less than 0 and greater than 1
- The model will not produce the probability of the dependent variable's occurrence that this thesis seeks to determine

Two models can overcome the disadvantages presented by the linear regression model; they are the probit (probability + unit) and logistical regression (logit) models. Probit and logit have become the standard method of analysis for situations where the outcome variable is dichotomous (Hosmer and Lemeshow 1989). Probit and logit better handle dichotomous variables by using an iterative optimization routine to maximize a log likelihood function (Mun 2012).

Most of the reviewed literature states that the probit and logit models yield similar results; the choice of probit versus logit depends largely on individual preference (University of California, Los Angeles: Statistical Consulting Group 2013b). Additionally, probit tends to be more popular in economic studies, while logit tends to be used in the health sciences (Methods Consultants of Ann Arbor 2013). Both logit and probit share the ability to handle dichotomous variables, and their outputs forecast the probability of an event's occurrence. Since their results are nearly indistinguishable and the area of this thesis is more aligned to economics than health sciences, the probit model was chosen to conduct the analysis.

The probit regression model is a type of maximum likelihood estimation technique that is used to forecast the likelihood of something occurring given one or more independent variables. In probit regression models the dependent variable can take on only two values, while the independent variables can be non-dichotomous. Once again, the probit model uses the independent variables listed in Table 19 and adds the dependent variables, cancelled and non-cancelled programs.

Because of the very small sample size, each variable is regressed against cancellations individually. The anticipated output was three-fold: (1) reinforcement of the statistically significant independent variables, (2) predicted probabilities of program cancellation resulting from changes in significant independent variables, and (3) graphical representation of the significant probit models. The mathematical expression of the probit model that estimates the probability of event Y_i occurring is expected to be in the form of Equation (13):

$$P(Y_i = 1) = \Phi(\beta X_i) \tag{13}$$

Where Φ is the cumulative density function of the standard normal distribution, β is the marginal effect of the explanatory variable on the dependent variable, and X_i represents the values of the explanatory variables of interest. Software packages such as STATA® (STATA® is a general purpose statistical software package developed by StataCorp) utilize maximum likelihood estimation and numerical optimization to approximate β (Koop 2012). A graphical representation of the probit model will be useful in interpreting the probability of cancellation given a certain explanatory variable value. Probit uses an

S-shaped normal distribution cumulative density function to estimate the curve of the data instead of a linear function (O'Halloran 2013). Figure 8 represents a notional curve fitting the data for a non-dichotomous independent variable and dichotomous dependent variable.

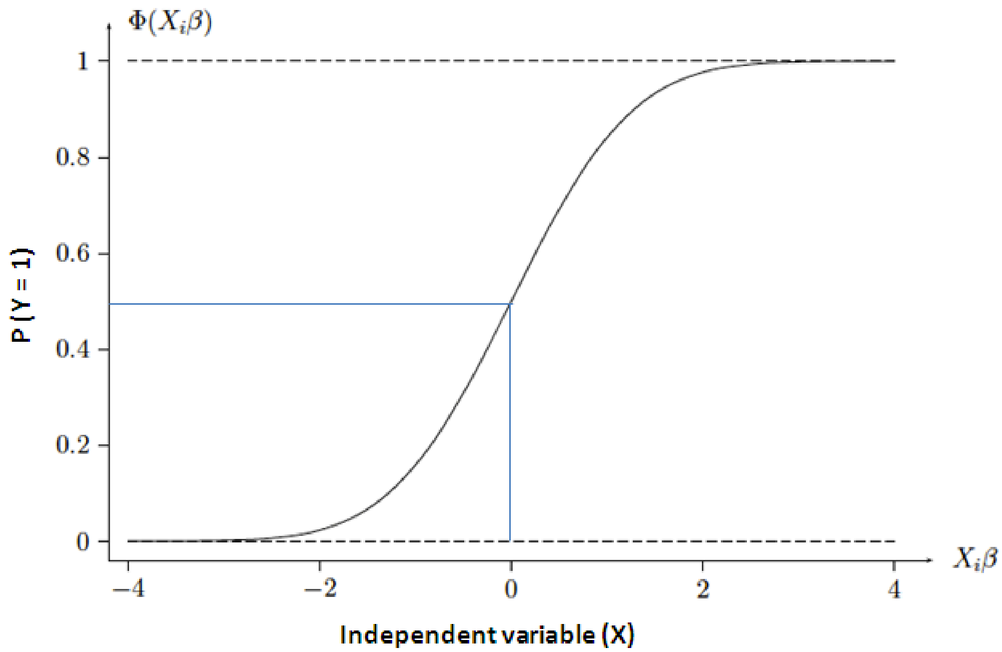

Figure 8. Graphical representation of a notional probit model
(From University of Zurich: Chair for Statistics and Empirical Economic Research, http://www.sts.uzh.ch/index.html).

The independent variable, X, in this case can take on any value between -4 and 4. Observations of the dependent variable, Y, take on only one of two values, 0 or 1. Probit fits an S-shaped curve to the data; this curve can be used to approximate the probability of the dependent outcome based on the independent variable. An example will help illustrate the utility of this model. To determine the value of the independent variable X at which the probability of event Y occurring is 0.50, i.e. P(Y=1) = 0.50, find 0.50 on the y-axis and connect a line from 0.50 to the curve. Next, connect a line from this point on the curve down until it intersects the x-axis. This occurs approximately when X = 0. This value is called the “point of equal opportunity” and is useful because for values of X greater than this point (all Xs > 0 in this example), there is a greater likelihood of one of the binary events occurring. This will be useful in determining the independent variable threshold where a program is more likely to be cancelled.

C. SUMMARY

This chapter provides the rationale and approach for analyzing the data using Mann Whitney and probit statistical techniques, and explains their relative effectiveness given the constraints of the data. The next chapter presents the results and the value of those results to program managers, systems engineers and decision makers. Chapter VI presents ideas for taking the results and creating tools and mechanisms to enhance effective decision making.

THIS PAGE INTENTIONALLY LEFT BLANK

V. RESULTS

This chapter presents the results of this research and discusses the insights gained from the results. First, Mann Whitney test results are presented that reveal differences in the key earned value metrics of cancelled and non-cancelled programs. Then, a probit regression model is developed that indicates the probability of cancellation based on certain earned value variables. The results of the Mann Whitney tests and the probit model are presented. Additionally, this chapter provides information regarding the relevancy and utility of the results for decision makers, systems engineers, and program managers.

A. MANN WHITNEY TESTING

1. Tests for Differentiating Cancelled and Non-Cancelled Programs with Earned Value Variables

Discovering a variable or multiple variables amongst the EV data that distinguish cancelled programs from non-cancelled programs is one of the main objectives of this thesis. Six main groups of EV variables were tested using Mann Whitney:

- Cost variance percentage
- Schedule variance percentage
- Cost growth (PM EAC)
- Cost growth (Contractor EAC)
- Differences between PM and contractor EAC
- Variance reset frequencies

Additionally, the timing of the first five groups of earned value metrics was tested. The first four groups were tested at the 25% and 50% completion points of the contract and the fifth group was tested for each of the first three quarters of the contract. The Mann Whitney tests were performed on these variables using an alpha of 10% as the rejection criteria. The results of each individual test are summarized in Table 20.

Variable tested	p-value	Null Hypothesis
Cost Variance %	0.0517*+	Reject
Cost Variance % at 25% completion	0.1626	Fail to reject
Cost Variance % at 50% completion	0.1927+	Fail to reject
Schedule Variance %	0.1592+	Fail to reject
Schedule Variance % at 25% completion	0.3734	Fail to reject
Schedule Variance % at 50% completion	0.5000	Fail to reject
Cost Growth (PM Estimate at Completion)	0.1136	Fail to reject
Cost Growth (PM Estimate at Completion) at 25% completion	0.0407**	Reject
Cost Growth (PM Estimate at Completion) at 50% completion	0.0822*	Reject
Cost Growth (Contractor Estimate at Completion)	0.0639*	Reject
Cost Growth (Contractor Estimate at Completion) at 25% completion	0.0533*	Reject
Cost Growth (Contractor Estimate at Completion) at 50% completion	0.0241**	Reject
Difference between Contractor and Program Manager's EAC	0.2816+	Fail to reject
Difference between Contractor and Program Manager's EAC (0-25%)	0.5000	Fail to reject
Difference between Contractor and Program Manager's EAC (26-50%)	0.2001	Fail to reject
Difference between Contractor and Program Manager's EAC (51-75%)	0.1362	Fail to reject
Cost Variance Reset Frequency	0.2474	Fail to reject
Schedule Variance Reset Frequency	0.4582	Fail to reject

*** $p < 0.01$, ** $p < 0.05$, *$p < 0.10$

+ indicates that first MW test returned N/A as a result for p-value and the test was re-run with swapped medians in the alternative hypothesis

Table 20. Summary of Mann Whitney results.

Note in Table 20 that four EV variables were accepted at alpha = 0.10 and two were significant at alpha = 0.05. Tests noted with a "+" sign were those Mann Whitney tests that were performed again because the median expected to be more unfavorable turned out to be less unfavorable. The one-tailed Mann Whitney tests were repeated with the medians swapped in the alternate hypothesis. These tests returned a p-value that could be analyzed. More details about these results are presented in the following sub-sections and in Appendix A.

a. Cost Variance Percentage

This study expects that cost variance percentages (CV%) would be more negative and unfavorable in cancelled programs than non-cancelled programs. The purpose of this Mann Whitney test was to determine if this was the case. Surprisingly, of the troubled programs sampled, non-cancelled programs' median CV% is significantly less (more negative and more unfavorable) than the sampled cancelled programs' median

CV%. This suggests that the sampled non-cancelled programs are more troubled than the cancelled programs, from a cost variance percentage perspective. This is counter-intuitive and unexpected. However, there is something to be learned from this result. Recall from Chapter III that cost variance percentage was used as the selection criteria for troubled programs. In order to reasonably compare the troubled programs, it is necessary to determine whether the second sampled group (non-cancelled troubled programs) is at least as troubled as the cancelled sample; if they aren't as "troubled" one could easily criticize the results on the basis that the second sample of non-cancelled programs were better performers and thus should not have been cancelled. Given the results of the previous test, a second Mann Whitney test was conducted to determine whether the non-cancelled programs have a significantly lower CV% (more unfavorable and more negative) than the cancelled programs. The results (p-value = 0.0518) suggest that this appears to be the case. Of the sampled troubled programs, non-cancelled programs' median CV% is less (more unfavorable) than the sampled cancelled programs' median CV%. This suggests that the non-cancelled programs are at least (if not more so) troubled than the cancelled programs, from a cost variance perspective. Since the non-cancelled programs chosen have performed worse than the cancelled programs from a median CV% perspective and survived, there may be some other earned value metric that indicates program cancellation.

The timing of cost variance percentages was investigated to see if CV% at certain contract completion points was indicative of program cancellation. This study expects that cancelled programs would have more unfavorable CV% at the 25% and 50% contract completion points than their non-cancelled counterparts. A Mann Whitney test (alpha = 0.10) was used to determine whether a difference in the samples' CV% at contract completion percentage levels of 25% and 50% (there was not enough data available at the 75% completion point for sufficient analysis) existed between the sampled cancelled and non-cancelled programs. For the first completion point tested, the average of the CV% within +/- 5% of 25% completion point was computed and used for the cost variance of each program. While the median CV% at the 25% contract completion point is more negative (and unfavorable) for cancelled programs, the p-value from this test result is greater than alpha = 0.10, suggesting CV% is not a significant indicator of cancellation at the 25% completion point. A similar test was conducted for CV% at the 50% completion point of the contract. The average of the cost variance percentages within +/- 5% of 50% completion point was computed and used for the CV% of each program. The median CV% at the 50% contract completion point is higher for cancelled programs than for non-cancelled programs thus a p-value is not available and the null hypothesis can not be rejected. There is insufficient evidence to conclude that the median CV% at the 50% completion point is more negative for the sampled cancelled programs. A second test was conducted to see if non-cancelled CV% at 50% was more unfavorable than cancelled CV% at 50%. Since the p-value of 0.1927 greatly exceeds the alpha = 0.10, the null hypothesis can not be rejected in this case either.

The results of this study of cost variance percentages suggest that, (1) non-cancelled programs had more unfavorable CV% than their cancelled counterparts, and (2) the timing of unfavorable CV% does not appear to have an effect on cancellation.

b. Schedule Variance Percentage

Attention was turned to testing other EV variables to see if there was a statistically significant discriminating variable of cancelled programs. The next variable tested was schedule variance percentage (SV%). It seems reasonable to expect that cancelled programs would have a harder time meeting task completion deadlines and schedule milestones on time. Therefore, SV% is expected to be worse for cancelled programs than non-cancelled programs. The results of the Mann Whitney test indicated otherwise. Similar to the CV% results, the median SV% of the non-cancelled programs was actually worse than the median SV% of cancelled programs. Because of this, there is no evidence that the median SV% of cancelled programs is more unfavorable than the median SV% for non-cancelled programs. Given these results, the test was run again to check to see if the non-cancelled programs had significantly worse SV% than cancelled programs. Test results revealed a p-value = 0.1592 that exceeded alpha = 0.10; these results indicate that the two samples were not statistically different from a schedule variance percentage perspective.

An investigation into the timing of schedule variance percentages examined if SV% at certain contract completion points was indicative of program cancellation. For the 25% completion point, the average of the SV% within +/- 5% of 25% completion point was used for the SV% of each program. Again, a Mann Whitney test was used to test whether the median SV% for cancelled programs was more unfavorable than their non cancelled counterparts at the 25% completion point. While the median SV% at the 25% contract completion point is lower for cancelled programs, the results (p-value = 0.3734) revealed there is insufficient evidence to show that cancelled programs have more unfavorable SV% than non-cancelled programs at the 25% completion point. A similar test was conducted for SV% at the 50% completion point of the contract. The test results (p-value = 0.5) revealed that the there is no evidence of more negative median SV% at the 50% completion point for the sampled cancelled programs. In fact, there is an equal likelihood that there is an effect as there is an appearance of an effect due to randomness.

The results of this schedule variance percentage study suggest that there is not a significant difference between the SV% or the timing of SV% in cancelled and non-cancelled programs. The study shifted its focus to investigating cost growth as a potential explanatory variable of cancellation.

c. Cost growth

Recall from Chapter II, that the earned value metric chosen to measure cost growth is estimate at completion and that estimates at completion come in two forms: program manager's estimate at completion and contractor's estimate at completion. Both the program manager's and contractor's estimates were tested separately since there is a difference between the estimates in most programs. This portion of the study investigates whether total cost growth over the life of the contract is worse in cancelled programs. Furthermore, it studies whether the timing of cost growth is worse in cancelled programs than their non-cancelled counterparts.

(1) Program Manager's Estimate at Completion. Since total cost growth is an unfavorable metric in the eyes of program stakeholders, this study

expects that cost growth in the form of PM EAC would be worse for cancelled programs. There is an observed difference in the medians of the two samples (89.9% vs. 45.8%) and it appears that cancelled programs have higher total cost growth than non-cancelled programs. However, with this data there is not compelling evidence to reject the null hypothesis in favor of the alternative. While the traditional interpretation of a p-value of the Mann Whitney results (p-value = 0.1136) suggest that the null hypothesis cannot be rejected at the 10% level of significance, this relatively low p-value may suggest an area for further investigation into the relation between total cost growth of the program manager's EAC and program cancellation.

Next, the study investigated the timing of cost growth in the form of the program manager's estimate at completion. It was expected that early cost growth and midpoint cost growth would be worse in cancelled programs than non-cancelled programs. A Mann Whitney test with alpha = 0.10 was used to determine whether a difference in the samples' cost growth at contract completion percentages of 25% and 50% (again, there was not enough data available at the 75% completion point for sufficient analysis) existed between the sampled cancelled and non-cancelled programs. For the two completion points tested, the average cost growth for completion points within +/- 5% of 25% and 50% completion points, respectively, were computed and used for the cost growth of each program. At the 25% completion point, there was an observed difference in the cost growth medians of the two samples (33.1% vs. 0.4%). It appears that cancelled programs have significantly higher cost growth than non-cancelled programs at the 25% completion point. Moreover, the Mann Whitney test results (p-value = 0.0407) indicate that there is evidence that the cost growth of the program manager's estimate at completion at the 25% contract completion point is greater for cancelled programs than non-cancelled programs.

Median cost growth continues to increase as the contracts reach their midpoint of completion. For non-cancelled programs, the median cost growth increased from 0.4% to 16.4%. Similarly, for cancelled programs, the median cost growth rose from 33.1% to 47.2%. This is expected as programs seldom recover from the inefficiencies that cause early cost growth. This is also consistent with Christensen's

(1994) findings that once a program is more than 15–20% complete it is highly unlikely that the cost overrun for a program will be less than the current cost overrun. Moreover, the Mann Whitney test revealed that there is evidence that the cost growth at the 50% completion level is higher for cancelled programs than non-cancelled programs. The utility of this finding is discussed later in this chapter.

(2) Contractor's Estimate at Completion. Recall from Chapter II that the contractor's estimate at completion is a measure of cost growth; if a difference exists between the contractor's and program manager's estimate it is typically due to use of a different index to calculate the estimate. Nevertheless, this study expects that the growth in the contractor's estimate, like the program manger's estimate, would be higher in cancelled programs than non-cancelled programs. The median total cost growth of cancelled and non-cancelled programs was compared and there is an observed difference between the two samples (91.7% vs. 32.3%). Again, it appears that cancelled programs have significantly higher total cost growth than non-cancelled programs over the life of the contract. The Mann Whitney test confirmed this expectation (p-value = 0.0639) and showed that there is statistically significant evidence that the total cost growth of the contractor's EAC is greater for cancelled programs than non-cancelled programs.

The study expects that the cost growth in the form of the contractor's EAC at the 25% and 50% contract completion point would be worse in cancelled programs than non-cancelled programs. For the two completion points tested, the average cost growth was calculated in the same manner as for PM EAC in the preceding test. As expected, the median cost growth of the cancelled programs is higher than the median cost growth of the non-cancelled programs at the 25% completion point (32.7% vs. 3.2%) and the Mann Whitney test confirms that these results are significant. It can be concluded that the cost growth in the form of the contractor's estimate at completion is greater for cancelled programs than non-cancelled programs at the 25% contract completion point. The median cost growth worsens as the contract progresses towards completion for both cancelled and non-cancelled programs, 52.8% vs. 20.3%, respectively. The Mann Whitney tests confirm that cost growth in the form of the contractor's estimate at completion is greater for cancelled programs than non-cancelled programs at the 50% contract completion point.

Expectedly, these results are similar to those obtained for the program manager's estimate suggesting a trend in cost growth in cancelled programs. Cost growth in the estimates at completion was the strongest statistical evidence found of a difference between cancelled and non-cancelled programs. The utility of this finding is discussed later in this chapter.

d. Differences between PM and Contractor EAC

The next variable tested was the difference between the program manager's EAC and the contractor's EAC for both cancelled and non-cancelled programs. Since the EACs are calculated differently, the study seeks to explore whether a potential widening gap can indicate cancellation. This study would expect wider gaps to be associated with cancelled programs. The rationale here is that wider gaps indicate more uncertainty in the estimates and uncertainty in growing costs tends to prove unnerving to decision makers. Recall that the difference between the two estimates is calculated based on the values reported in each program's cost performance report in the DAES report and then normalized using the procedure detailed in Chapter III. The median normalized EAC difference was calculated for each program. Then the median EAC difference of cancelled programs was compared to the median EAC difference of the non-cancelled programs. Unexpectedly, the median EAC difference was actually greater for non-cancelled programs than cancelled programs. A second test was conducted to see if non-cancelled EAC difference was statistically worse than cancelled EAC difference; it was not. Therefore, there is no evidence that there is a difference between the median EAC difference for the sampled cancelled and non-cancelled programs.

The difference between the program manager's EAC and the contractor's EAC over time was also investigated. The normalized differences of the estimates were investigated for three periods (0–25%, 26–50%, 51–75%). The study expected that cancelled programs would experience greater gaps in the estimates at all phases of the contract. The p-values of the first two Mann Whitney tests (for the periods 0–25% and 26–50%) greatly exceeded the alpha of 0.10 and therefore there was no evidence that EAC differences were related to cancellation at these phases of the contract. The third test for the period 51–75% resulted in a p-value of 0.1362. The p-value slightly exceeds the alpha of 0.10, suggesting that there is some evidence that the difference between the program manager's and contractor's estimate at completion is greater for cancelled programs than non-cancelled programs at the 51–75% completion level. This relatively low p-value may suggest an area for further investigation.

2. Effect of Cost and Schedule Variance Resets on Program Cancellation

One of the research questions involves investigating whether resetting cost variance or schedule variance increases the likelihood of program survival. The small sample sizes made this question difficult to answer, especially considering that the range of frequencies was zero to three. The schedule and cost variance frequencies were counted for each of the 16 programs. It was initially observed that on average the cancelled programs experienced 0.75 cost variance resets and 0.50 schedule variance resets, whereas the non-cancelled programs experienced 0.50 cost variance resets and 0.63 schedule variance resets. Since the average number of cost variance resets for cancelled programs exceeds the average number of cost variance resets for non-cancelled programs, one can reasonably conclude that resets of cost variances do not decrease likelihood of cancellation. Similarly, since the average number of schedule variance resets in cancelled programs is only slightly lower than those in non-cancelled programs, this cursory study of mean frequency of schedule variance resets does not appear to suggest an association with cancellation.

Nevertheless, Mann Whitney was used to analyze whether the median cost and schedule reset frequencies were associated more with cancelled or non-cancelled programs. If reset frequency is higher in cancelled programs, then they may be a leading indicator to predict cancellation. If resets saved programs, then their frequency would be expected to be higher in the non-cancelled troubled programs that survived. Neither conclusion can be drawn however since the p-value greatly exceeds the alpha on the tests conducted.

Based on the data set, cancelled programs do not appear to have more cost or schedule variance resets than their non-cancelled counterparts. Further analysis could be conducted on the size and timing of the resets and their relationship to program cancellation or survival.

3. Interpretation of the Mann Whitney Results

Of the sampled troubled programs, the cancelled programs have higher total cost growth over the life of the contract based on contractor EAC and program manager EAC. Additionally, the analysis suggests that cost growth at the 25% completion level and again at the 50% completion level (for both the program manager and contractor's estimate at completion) was greater for cancelled programs than non-cancelled programs. While this seems intuitive, this study found affirmative evidence to confirm the expectation that cancelled programs tend to have greater cost growth at the beginning and midpoint levels of contract completion. Additionally, the increasing median cost growth from 25% to 50% completion point is valuable to program managers and decision makers because it confirms the findings that cost growth worsens (and rarely improves) over time. Christensen (1994) notes that given that a contract is more than fifteen percent complete, the overrun at completion will not be less than the overrun incurred to date; and the percent overrun at completion will be greater than the percent overrun incurred to date. These findings serve as another foreboding example of the consequences of excessive cost growth.

What remains puzzling is the finding that non-cancelled programs experienced worse (more negative) cost variance percentages than the cancelled programs, yet their median contractor and PM EAC growth (which presumably uses CV % in its calculus) was more favorable than their cancelled counterparts. The expectation is that those programs with worse cost variance would have greater growth in their EAC. This could be a result of limited sample size, or a result of the fact that the medians are not sufficiently explanatory, or caused by the use of different indices to calculate contractor and PM EAC. As explained in Chapter II, the indices used by programs to calculate EAC vary and are unknown for the programs examined in this research; a deeper investigation into the indices for each respective program could be conducted to analyze if optimistic, pessimistic or realistic estimations were associated with cancellations. The point here, however, is that there are numerous EAC formulas that can be used by the contractor and program offices, each with their own assumptions and motivations for use. Chapter II goes into detail regarding the different methods for calculating EAC. Programs have the latitude to lower estimate at completions by using different performance indices, none of which may be indicative of future cost performance of the contract (Christensen 1994 and U.S. Government Accountability Office 2009). This may divorce EAC from the more negative performance indices and may present a lower than realistic EAC. While this thesis was able to draw conclusions about the higher growth of cancelled programs using these deterministic EACs, they are sub-optimal in providing value to program managers and decision makers who should have access to quality (risk informed) estimates to enhance their decision making. The Navy's A-12 program is the classic example of the problem with deterministic and non-uniform EACs. When Secretary Cheney cancelled the program in 1991, he lamented about the many inconsistent EACs (Morrison 1991). This can be improved.

Since this thesis highlights the significance of cost growth as measure by EAC at all phases of the contract, it seems reasonable to focus on making EAC as consistent and realistic as possible. A reasonable approach to standardizing the EAC calculation and including an associated level of certainty would seemingly reduce the inconsistency associated with programs' EACs, simplify the process for program managers, and provide program managers and senior level decision makers with control tools and higher quality estimates for evaluation within and across programs. Chapter VI discusses a method to improve the quality of information of the EAC metric to all program stakeholders.

The findings of the Mann Whitney analysis point to cost growth as the greatest differentiating earned value factor of cancelled programs. The thesis turns its focus to developing a model that is capable of confirming these differentiating earned value metrics and forecasting the probability of program cancellation based on earned value information.

B. PROBIT MODELS

Once explanatory variables were identified using Mann Whitney testing, the research shifted focus to developing a model to forecast the probability of cancellation based on EV information. Multiple probit models were developed using the analytical approach detailed in Chapter IV. The following sub-sections highlight both the results of that effort and the model’s value to program managers, systems engineers and decision makers. Additional probit test results details are presented in Appendix B.

1. Probit Models-– All Programs in Study

The same EV variables of all 16 programs used in the Mann Whitney tests were utilized in the development of a probit model. Because of the small sample size each individual variable was run as a separate regression to avoid losing degrees of freedom.

The anticipated output of each probit model was three-fold: (1) reinforcement of the statistically significant independent variables using parametric estimators, (2) predicted probabilities of program cancellation resulting from changes in significant independent variables, and (3) graphical representation of the significant probit models. For p-values less than alpha = 0.10, the variable was considered significant. Recall that the p-value is the probability that the coefficient in Equation (13) is actually equal to zero, and thus a small p-value indicates that the coefficient is likely non-zero. A p-value of less than 0.10 indicates that there is less than a 10% chance the model’s results for the particular variable would be the result of random behavior. The complement is true as well; the model is 90% confident that the independent variable is having an effect on the dependent variable. Table 21 contains the p-value and coefficient[13] results from the probit tests.

[13] Recall from Chapter IV that probit coefficients are transformed by the cumulative density function and cannot be interpreted in the same way as OLS regression coefficients.

Variables	P-value	Coefficient
CV%	0.151	0.186
CV% at 25% completion	0.186	-0.067
CV% at 50% completion	0.791	0.016
CV reset frequency	0.473	0.326
SV%	0.796	-0.019
SV% at 25% completion	0.764	-0.012
SV% at 50% completion	0.373	-0.066
SV reset frequency	0.751	-0.127
Cost growth (PMEAC)	0.256	0.624
PMEAC at 25% completion	0.090	4.029*
PMEAC at 50% completion	0.251	1.446
Cost growth (Contractor EAC)	0.102	1.449
Contractor EAC @ 25% completion	0.095	4.454*
Contractor EAC @ 50% completion	0.163	1.757
Diff. b/w Contr. EAC & PMEAC	0.577	2.290
Diff. b/w Contr. EAC & PMEAC (0-25%)	0.706	2.477
Diff. b/w Contr. EAC & PMEAC (26-50%)	0.440	-6.431
Diff. b/w Contr. EAC & PMEAC (51-75%)	0.633	2.225
*** p<0.01, ** p<0.05, * p<0.10		

Table 21. Probit results.

Note from Table 21 that the only two variables for which the p-values are less than 0.10 are the growth of the program manager's estimate at completion at the 25% completion point and the growth of the contractor's EAC at the 25% completion point. These significant findings for these two variables reinforce the results of the Mann Whitney tests, which also found these two variables to be significant. These two variables were used to develop individual probit models whose chief utility is its ability to indicate the probability of cancellation given variable input data. The high correlation of the two estimates at completion prevents combining them into one regression model.

Figure 9 is the graphical representation of the first model utilizing Cost Growth PM EAC at 25% complete. This graphical representation of the model depicts the probability of cancellation given a certain level of cost growth in the program manager's estimate at completion at the 25% completion level of the contract.

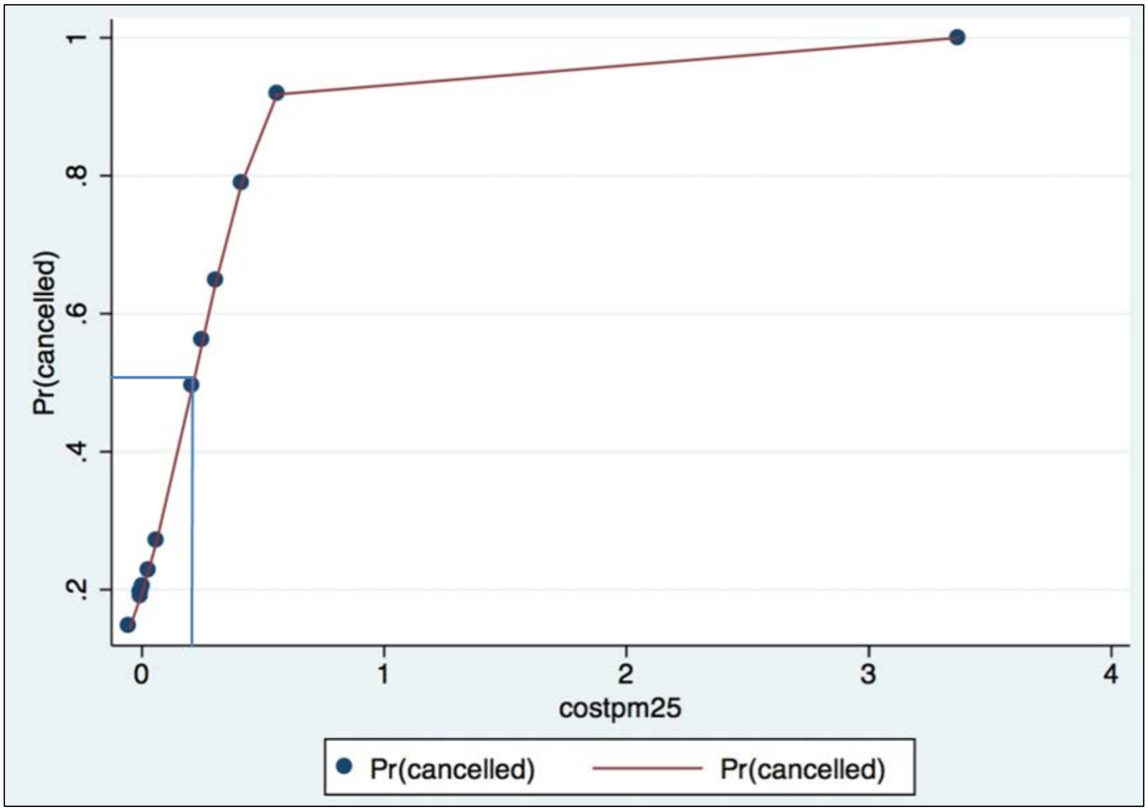

Figure 9. Cost growth PM EAC at 25% completion probit model[14].

From Figure 9, for cost growth greater than 20%, the probability of cancellation is greater than 50%. Recall from Chapter IV that this point is called the point of equal opportunity. This point is meaningful to program managers, systems engineers, and decision makers because it is the cost growth threshold where a program is more likely to be cancelled than not cancelled. Probit provides another useful tool for decision makers; the estimated marginal effect is the average slope of the S-curve. For this model, the estimated marginal effect is 1.10. On average, when cost growth increases by 1%, a decision maker can expect an approximate 1.1% increase in likelihood of cancellation. This is a valuable rule of thumb for decision makers involved with a program to get a sense of the effects of increasing cost growth in the early stages of a contract.

The second model was developed using the Cost Growth Contractor EAC at 25% and is very similar to the probit model for Cost Growth PM EAC at 25%. Figure 10 is a

[14] The reader will note that x-axis units are (x100%)

graphical representation of the probit model using Cost Growth Contractor EAC at 25% completion point as the independent variable.

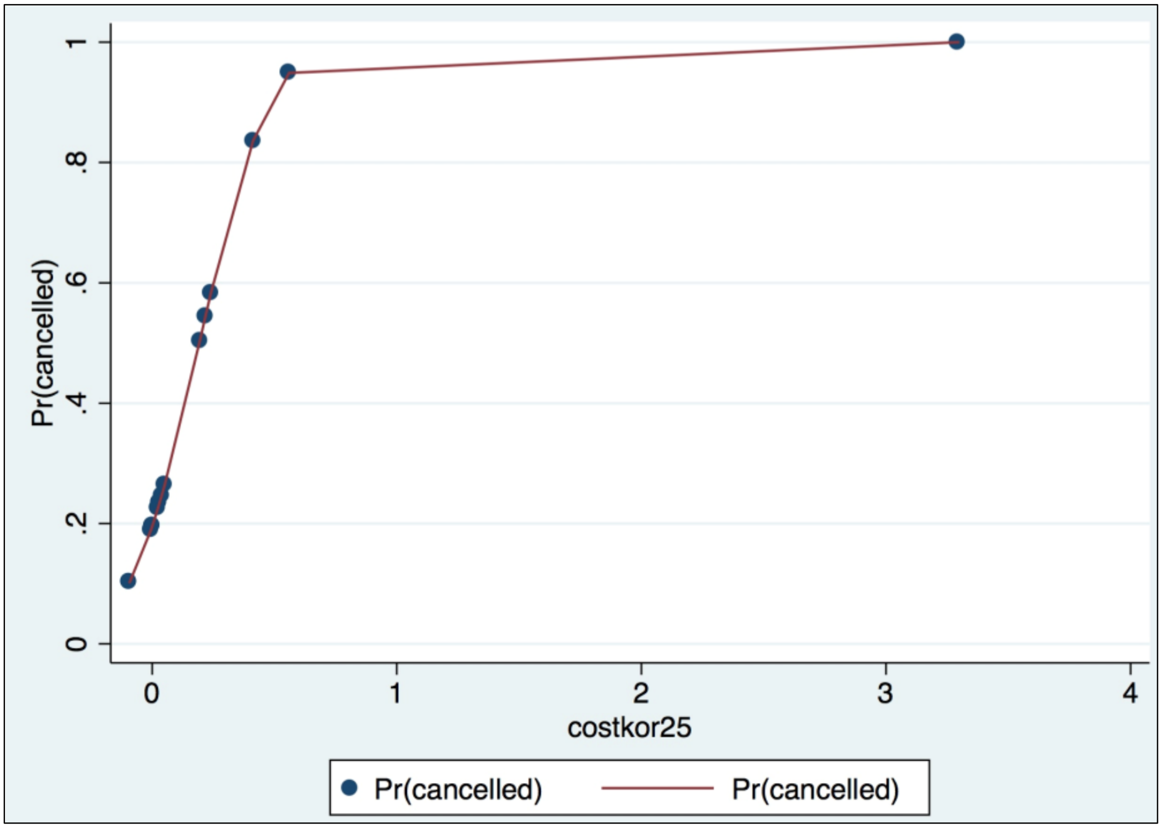

Figure 10. Cost growth contractor EAC at 25% completion probit model.

Again, it is evident that this graphical model is useful as a tool to indicate cancellation based on a corresponding contractor EAC at 25% contract completion point. Visually, it is difficult to discern an appreciable difference between the two models; however, the cost growth contractor EAC at 25% maintains a slightly higher average slope. The estimated marginal effect for this model is 1.21, which means that on average when cost growth increases by 1%, a decision maker can expect a corresponding 1.21% increase in likelihood of cancellation. Nevertheless, either graph can be used by program managers, decision makers, and systems engineers as a quick reference tool for understanding the effect of early cost growth on the probability of program cancellation. These significant probit results in conjunction with the Mann Whitney test results reinforce the notion that early cost growth is significant in forecasting likelihood of program cancellation.

2. Revised Probit Models

There is a potential outlier in the probit model in Figures 9 and 10. In order to test the sensitivity of the models and to ensure that the significance was not being driven by this sole observation, the probit modeling was repeated for the data sets excluding the Crusader program whose cost growth (both contractor and program manager's EAC) was roughly 330% at the 25% contract completion point. Table 22 shows the comparison of the results between the two models with Crusader and the two models without Crusader.

Variables	With Crusader		Without Crusader	
	P-value	Coefficient	P-value	Coefficient
CV%	0.151	0.186	0.193	0.168
CV% at 25% completion	0.186	-0.067	0.195	-0.065
CV% at 50% completion	0.791	0.016	0.864	0.010
CV reset frequency	0.473	0.326	0.561	0.266
SV%	0.796	-0.019	0.651	-0.034
SV% at 25% completion	0.764	-0.012	0.805	-0.009
SV% at 50% completion	0.373	-0.066	0.586	-0.039
SV reset frequency	0.751	-0.127	0.637	-0.193
Cost growth (PMEAC)	0.256	0.624	0.428	0.544
PMEAC at 25% completion	0.090	4.029*	0.090	4.029*
PMEAC at 50% completion	0.251	1.446	0.309	1.411
Cost growth (Contractor EAC)	0.102	1.449	0.102	1.449
Contractor EAC @ 25% completion	0.095	4.454*	0.095	4.454*
Contractor EAC @ 50% completion	0.163	1.757	0.163	1.757
Diff. b/w Contr. EAC & PMEAC	0.577	2.290	0.511	2.776
Diff. b/w Contr. EAC & PMEAC (0-25%)	0.706	2.477	0.706	2.477
Diff. b/w Contr. EAC & PMEAC (26-50%)	0.440	-6.431	0.608	-4.379
Diff. b/w Contr. EAC & PMEAC (51-75%)	0.633	2.225	0.504	3.275
*** p<0.01, ** p<0.05, * p<0.10				

Table 22. Initial probit model vs. revised probit model results.

Notably, there is not much difference between the model with and without the Crusader program data. PM EAC at 25% completion and Contractor EAC at 25%

completion remain the only significant variables whose p-values are less than alpha = 0.10. This model was largely insensitive to excluding the outlier.

Figures 11 and 12 are the graphical representations of the revised probit models utilizing Cost Growth PM EAC at 25% completion and Cost Growth Contractor EAC at 25% completion as the model's independent variables, respectively.

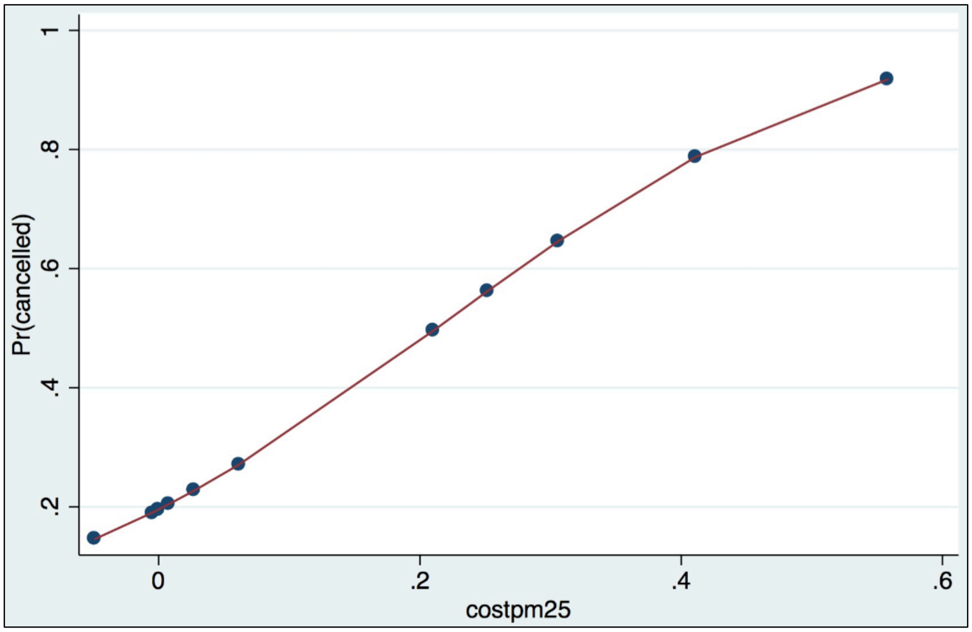

Figure 11. Cost growth PM EAC at 25% completion probit model (without Crusader).

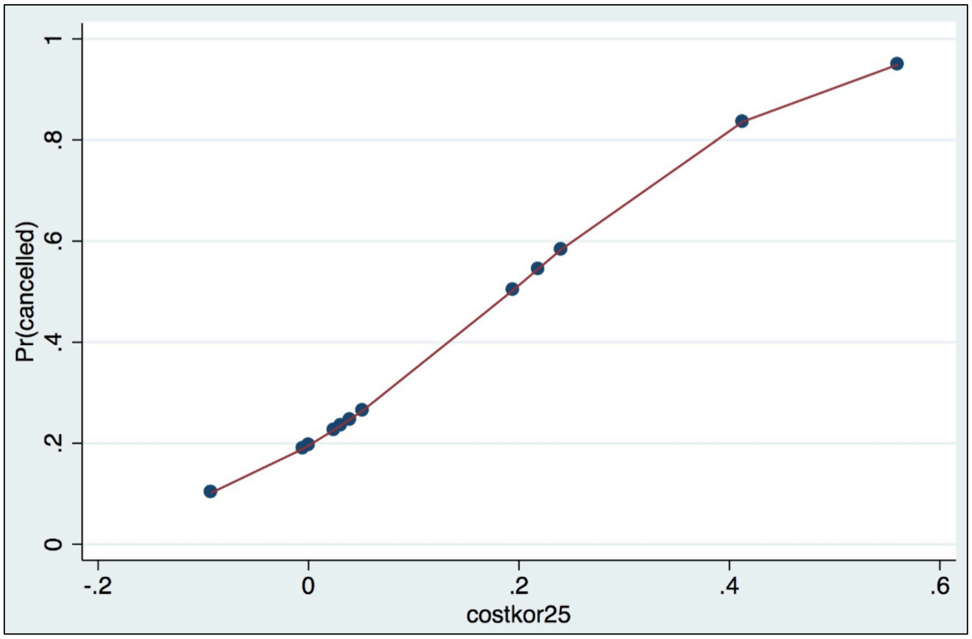

Figure 12. Cost growth contractor EAC at 25% completion probit model (without Crusader).

Figures 11 and 12 offer similar graphical representations as their predecessors. Although visually the average slopes of the S-curves appear to be shallower, they are nearly identical to those in the original probit model graphs. The estimated marginal effect for Cost Growth PM EAC at 25% completion is 1.19, up slightly from 1.10 in the original model. Similarly, the estimated marginal effect of the cost growth contractor EAC at 25% completion increased slightly from 1.21 to 1.30. On average, when cost growth increases by 1%, a decision maker can expect a corresponding 1.19% or 1.30% increase in likelihood of cancellation based on these marginal effects.

C. SUMMARY

The data indicate that of the troubled programs sampled, cancelled programs do not have a lower or more unfavorable cost variance percentage than non-cancelled programs. However, the Mann Whitney test results have shown that for the sampled programs, the cancelled programs have higher total cost growth over the life of the contract based on contractor EAC and program manager EAC. More specifically, the Mann Whitney analysis suggests that cost growth at the 25% and 50% completion levels (for both the program manager and contractor's estimate at completion) is greater for cancelled programs than non-cancelled programs. Moreover, probit modeling confirms that the significant variables are cost growth at 25% completion for both program manager and contractor EAC. Taken together, both the Mann Whitney test results and probit model results suggest that there is reasonably strong evidence that early cost growth in a program is an indicator of program cancellation. This information serves as a red flag to all program stakeholders and confirms a widely held notion that early cost growth in a program is indicative of program cancellation.

These findings have drawn attention to the importance of realistic estimates of the cost of a contract at completion. Recall from Chapter II that most EACs are calculated deterministically and reported in DAES reports and SARs to Congress. It stands to reason that while EAC alone is not the sole metric used in a cancellation decision, it does play a part. It seems reasonable to contend that the EAC is a critical metric for used by decision makers. Assuming that this is the case, the accuracy and realism of these estimates is particularly important. The following chapter discusses using more of the information from the Comprehensive EAC and probabilistic modeling of the monthly EACs as reasonable approaches to improving the information provided to decision makers.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. PRACTICAL IMPLICATIONS FOR DEFENSE ACQUISITION

A. OVERVIEW

Both the Mann Whitney and probit results highlight the importance of early cost growth as an indicator of cancelled programs. Moreover, the results show that early cost growth can be used to forecast the likelihood of program cancellation. Two probit models effectively forecast the likelihood of cancellation based on EAC growth at the 25% completion point of a contract. Because so much of this thesis' findings are predicated on the EAC, it follows then, that in order to maximize its utility to senior officials making program cancellation decisions, the EAC should be as realistic as possible while minimizing additional effort and cost. This chapter identifies some key shortfalls of current EAC reporting. It then provides some suggestions to remedy the shortfalls, thereby improving the quality and utility of the information provided to program managers for decision making and to defense officials and Congress for making key budget choices and possible cancellation decisions.

B. SHORTFALLS OF CURRENT ESTIMATES AT COMPLETION

The EAC metric is important for two reasons. First, it provides the program manager a final estimate of costs to ensure that there are sufficient funds programmed and authorized to cover the estimated costs. Second, it informs senior government executives about the total estimated cost of the programs in the DoD portfolio for their use in budgeting and decision-making. Ostensibly, it is reasonable to imagine that all stakeholders would prefer to have the most realistic information available upon which to base their decisions. The EAC in the current reporting system may not be maximizing the quality of information provided to program managers and decision makers. The three key shortfalls of current EACs are: (1) the use of different performance indices in the creation of the best and worst case EACs makes comparison within and between programs challenging, (2) the best and worst case EACs reported in DAES are deterministic estimates that do not provide the decision maker with the likelihood of exceeding the estimate, and (3) the current comprehensive EAC, while the result of probabilistic modeling, officially reports only the “most likely” EAC and is produced only once per year.

Before discussing the shortfalls further, it is worth reemphasizing how EACs are reported to decision makers. Recall from Chapter II that the two main reporting mechanisms to Congress, PARCA, and DoD officials of the latest estimates of cost, schedule and performance status are the Selected Acquisition Reports (SARs) and Defense Acquisition Executive Summaries (DAES). Two EACs are reported in the SARs: (1) Contractor EAC and (2) Program Manager EAC. This thesis assumes that these SAR EACs are the “most likely” annual CEAC estimates discussed in Chapter II. These are the only probabilistic estimates this research has discovered that are contained in official reporting[15]. Multiple EACs are reported in the DAES reports: (1) Program Manager Best Case, Worst Case and Most Likely EACs, and (2) Contractor’s Best Case, Worst Case and Most Likely EACs. The “most likely” EAC is again assumed to be the CEAC and the best case and worst case EACs are deterministic, i.e., point estimates calculated using various performance indices[16]. One can easily imagine that these numerous estimates that apply different performance indices in their calculation can make comparisons of EACs both within and across programs difficult.

While the DoD provides guidelines for which EV performance indices to use in the calculation of best and worst case EAC deterministic calculations (U.S. Department of Defense, Defense Contract Management Agency 2006 and Defense Acquisition University 2012a), they remain guidelines and not strict requirements. The contractors and program managers have the latitude to use different formulas for the contract’s best and worst case EAC calculations. Recall from Chapter II, the range of EACs is typically calculated using two different indices, the EAC CPI and the EAC composite, to calculate the best and worst case EACs, respectively. If contractors and PMs use different performance indices to calculate their floor and ceiling EACs, this can cause confusion within the program. While this study could not show a strong significant relationship

[15] Recall from Chapter II that neither the statute for SARs nor the SAR Data Entry Instructions state which EAC is used in SAR reporting.

[16] These calculations are shown in Chapter II.

between the difference in EACs and cancellation, there was some evidence at the 51–75% completion level that the difference between the PM and contractor EACs was greater for cancelled programs. This difference in the estimate is useful for PMs and decision makers since it suggests that uncertainty in estimation can increase likelihood of cancellation and could be investigated further in future research. Similarly, MDAPs using different formulas to calculate these EACs make comparisons between programs problematic. This inconsistency in estimates may not bode well for troubled programs. A program could easily use a performance index that would improve its relative standing amongst other programs using a more favorable, but less realistic performance metric to calculate its EAC (Christensen 1999 and U.S. Government Accountability Office 2009).

To understand the future cost of a system, the decision maker needs to know both the forecasted cost and the likelihood of the program exceeding that cost. Additionally, decision makers should understand the program’s uncertainties and how they are driving costs. Neither of these can be provided by the currently reported best and worst case EACs. The existing best and worst case EACs maintained in formal reports make no mention of the level of certainty (or uncertainty) of the estimate nor the likelihood of exceeding it. This is because these EAC calculations use past cost and schedule performance to predict future costs and implicitly assume that the future performance will be identical to the past. Because EACs are calculated in this way, it is widely recognized that earned value management cannot provide exact estimates at completion with associated levels of certainty. If risk-related estimates are required to be reported to decision makers, cost estimating and risk management techniques are needed (Defense Acquisition University 2009a). The best and worst case estimates do provide a range of estimates for decision makers, but these point estimates, either low or high, share a common feature: they lack information. The estimates do not provide the decision maker with any insight into the underlying uncertainty in the estimate or the likelihood of exceeding that estimate.

To account for this shortfall in reporting uncertainty along with the best and worst EACs, the DoD has offered guidance in documents such as Defense Contract

Management Agency's (DCMA's)[17] *Earned Value Management System (EVMS) Program Analysis Pamphlet (PAP)* (U.S. Department of Defense, Defense Contract Management Agency 2012) and *Examining the Comprehensive Estimate at Completion (CEAC)* (Makielski 2009) for calculating an annual Comprehensive EAC. The annual CEAC is calculated using "detailed estimating methods similar to those used in the proposal process[18]" and "it should be the most likely estimate with an expected probability of occurring of 50%" (Makielski 2009).[19] This CEAC seemingly remedies the shortfall of not reporting the associated uncertainty with the estimate; however it only reports the most likely case in the SARs and DAES. The shortfalls of the CEAC are that, (1) the complete CEAC risk information is not reported to decision makers in formal reports, and (2) it is produced annually, so it cannot be used as a real time tool for program managers to assess monthly performance.

In summary, these shortfalls involve under-reporting of available probabilistic EAC information. The following section discusses ways to use an existing CEAC, the Vargas EAC[20], and a combination of probabilistically determined EACs with the findings from this report's analysis to improve the quality of information provided to program managers and senior level DoD executives.

C. RECOMMENDED USES OF PROBABILISTIC EAC FOR PROGRAM MANAGERS AND DECISION MAKERS

The DoD values the use of cost uncertainty analysis in establishing probabilistic cost estimates that include the risks of exceeding them. The *GAO Cost Estimating and Assessment Guide* states, "the EAC must reflect the degree of uncertainty, so

[17] DCMA is the DoD's subject matter expert for EVMS.

[18] During the proposal process, cost estimators use cost uncertainty analysis to develop initial cost estimates to compare the estimates of competitors and to ensure the government obtains a fair price for the developed system and services provided (Garvey 2000). Additional details of how the CEAC is computed are contained in Chapter II of this thesis.

[19] The governing documents do not provide more detail about how CEAC should be calculated. This thesis assumes that the CEAC referenced in these documents is determined probabilistically using cost uncertainty analysis methods like those detailed in Garvey's *Probability Methods for Uncertainty Analysis* (Garvey 2000) and the GAO's *Cost Estimating and Assessment Guide* (U.S. Government Accountability Office 2009).

[20] The Vargas method for calculating probabilistic EAC is detailed in Chapter II.

that a level of confidence and risk can be given along with the point estimate" (U.S. Government Accountability Office 2009). Deterministic manipulation of EV data alone cannot provide these levels of confidence or the probability of exceeding the forecasted EAC. Moreover, cost uncertainty analysis is statutorily required. Both WSARA (U.S. Congress 2009) and the CEAC development documents (Makielski 2009) require its use.[21]. What remains puzzling is the continued use of deterministic best case and worst case EACs in formal reporting. This thesis makes two suggestions to improve the quality of EAC information provided to decision makers: (1) report the existing probabilistically determined CEAC along with more of its level of certainty information in the SAR, and (2) use the Vargas method for calculating a probabilistic EAC for DAES reporting and as a program management assessment tool. Finally, this section concludes by providing a recommendation to use probabilistic EAC information in conjunction with this thesis' probit model as a tool to forecast the likelihood of program cancellation. These next sections provide details on implementing these suggestions to remedy the shortfalls discussed in the preceding section of this report and improve the realism and utility of EAC information.

1. Use Probabilistic CEACs in Official Reporting

Chapter II describes the process for creating the CEAC—an annual probabilistic EAC that contains associated risk of overrun along with the estimate. This thesis suggests using the results of the CEAC to officially report EAC values that correspond to a 90% likelihood of overrun, a 50% likelihood of overrun, and a 10% likelihood of overrun obtained from the CEAC S-curve like the notional example that is presented in Figure 13.

[21] WSARA also requires disclosure of confidence level for baseline estimates for MDAPs, the rationale for selecting such confidence level, and justification if the level chosen is less than 80% (U.S. Congress 2009).

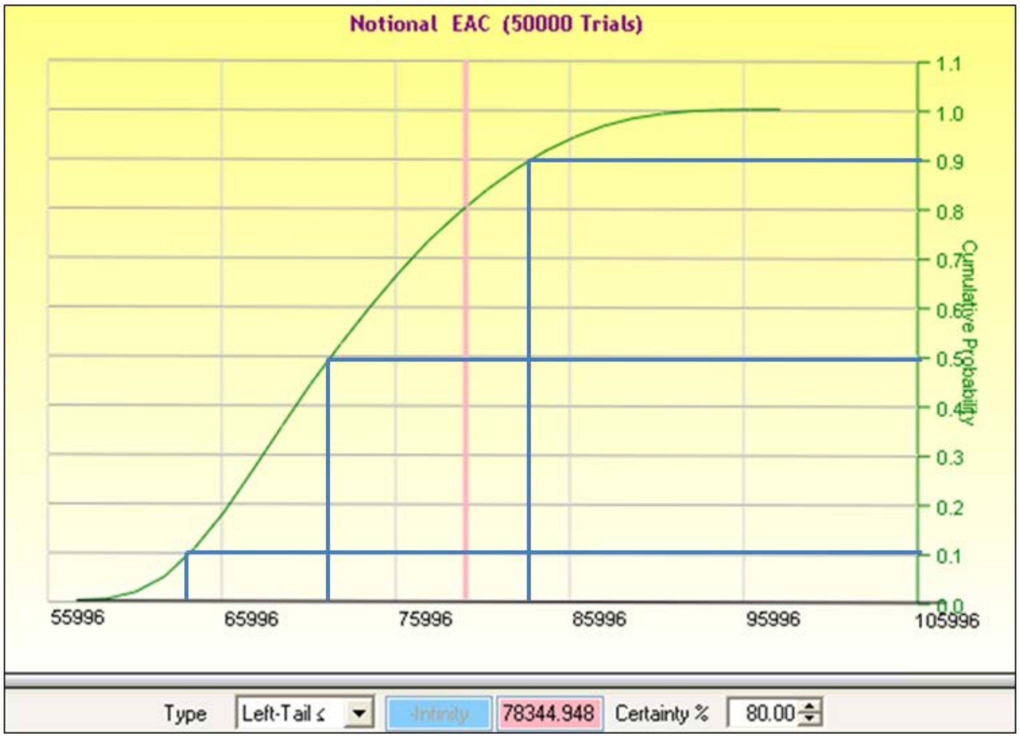

Figure 13. Probabilistic EACs with corresponding risk levels.

The valuable CEAC information is already available, so it is reasonable to include it in reports to Congress and senior level DoD decision makers. To standardize MDAPs' EAC reporting, DoD could add language to the DFAR governing EV and the EVMIG that requires contractors and programs to use a certain level of risk in their reporting for each EAC in the SAR. The benefit of this additional information is that Congress would now have a range of probabilistically determined EACs with associated levels of overrun risk instead of having only the "most likely" PM EAC and Contractor EAC to make cost growth assessments within programs and comparisons between programs. While this suggestion is not a major change, it would improve the quality of information provided to Congress in the SARs.

2. Vargas EAC as Program Management Assessment Tool

One of the shortfalls of the CEAC is that it is only produced annually and cannot provide program managers with current information based on the contract's monthly performance. Ricardo Vargas (2009) has proposed using the cost uncertainty analysis process to provide probabilistic EAC updates that contain overrun risk based on the monthly contractor reports. The primary benefits of this method are: (1) it provides monthly probabilistic estimates that can replace the deterministic estimates in the monthly DAES report, (2) it helps alleviate the ambiguity associated with the use of different performance indices, and (3) it provides the program manager with monthly probabilistic information that can be used as an assessment tool to monitor cost growth.

The novelty of Vargas' approach is its simplicity in applying existing monthly EV data in the CPRs to establish the parameters of the triangular distributions for the cost of each element; there is no need to go through the extensive process of establishing a distribution for each cost account as was done in the initial cost estimate or in the annual CEAC. A brief summary of the method is valuable to understanding its simplicity and usefulness. Recall from Chapter II that EACs can be calculated using numerous indices to forecast the cost of authorized work remaining. Vargas suggests that the three main performance indices used for deterministic EAC projections— constant (equal to 1), CPI, and SCI— be used as the parameters for a triangular distribution for each element's EAC. Use of the constant index, the CPI, and the SCI results in three different deterministic values for EAC—the best case, the most likely case, and the worst case, respectively. Furthermore, he proposes substituting these EACs for the minimum, most likely, and maximum values of the triangular distribution of the respective cost element as illustrated in Figure 14.

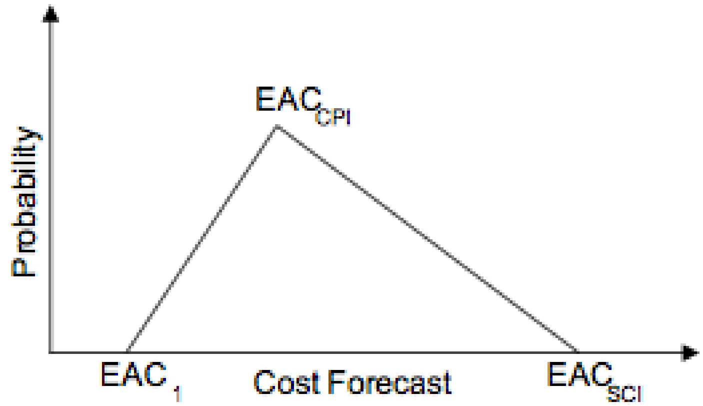

Figure 14. Triangular probability density function for WBS element EAC (from Vargas 2009).

The distributions are then used in a Monte Carlo simulation that sums the individual task's probabilistic EACs over repeated trials to generate a cumulative distribution function (CDF) (similar to the methodology used to calculate the CEAC detailed in Chapter II). Figure 15 shows the CDF or S-curve that would result for a hypothetical program.

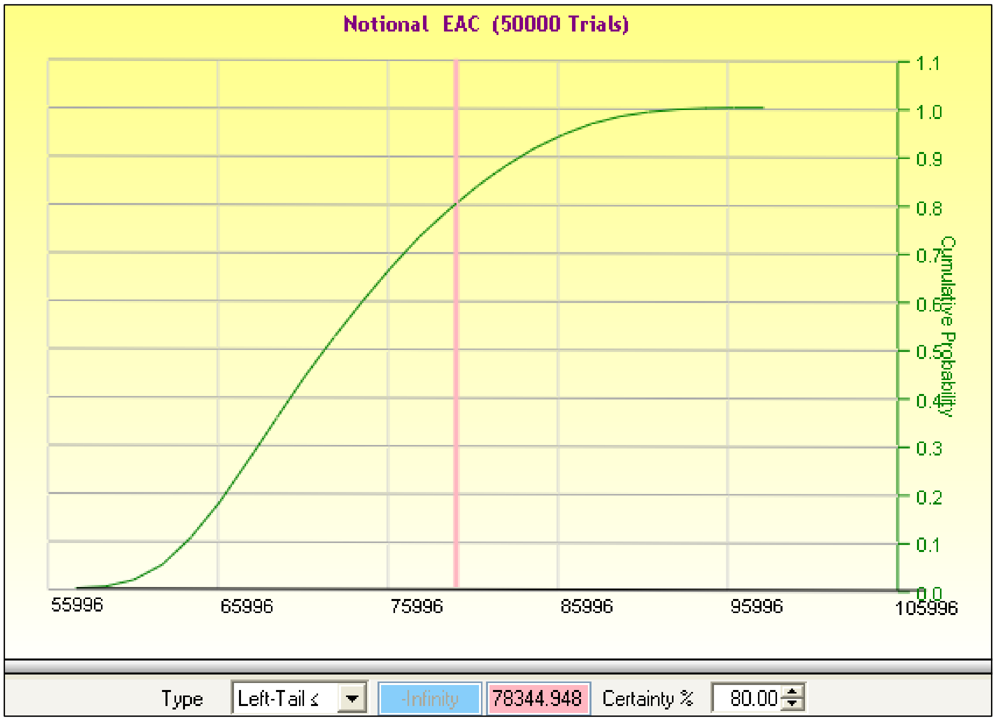

Figure 15. Cumulative density function for total EAC at 80% level of certainty.

The CDF (or S-curve) in Figure 15 shows the range of possible values for the total system cost, in this case $55,996 to $95,996. From the S-curve the program manager can easily determine the risk of exceeding any value between those two numbers. In Figure 15 there is an 80% chance that total cost will be less than $78,345 (as indicated by the pink line) or alternatively, the risk of exceeding $78,345 is 20%. Thus, the S-curve information can be used to determine the risk that the total cost will exceed a given budget, or the cost risk of the given budget.

Once the initial simulation model using the WBS cost elements has been created, it can easily be updated on a monthly basis using the EV data reported by the contractor. These probabilistic estimates can replace the deterministic estimates in the DAES reports. Because this method uses monthly performance data in its EAC calculus, program managers could use this CDF to compute the best case, worst case, and most likely EACs (like those listed in Table 23) on a monthly basis for inclusion into the DAES reports.

Level of Certainty	Upper EAC Bound
10%	$ 62,595.38
20%	$ 64,886.94
30%	$ 66,805.29
40%	$ 68,693.62
50%	$ 70,677.31
60%	$ 72,815.83
70%	$ 75,299.30
80%	$ 78,344.95
90%	$ 82,213.30

Table 23. EAC with levels of certainty quick reference table.

The Vargas method helps alleviate the ambiguity associated with the use of different performance indices. Since the three indices are accounted for in the underlying distributions for all programs, the result facilitates both comparisons within and between programs. Moreover, to standardize MDAPs' EAC reporting in DAES, DoD could add language to the DAES governing documents that require contractors and programs to use a certain level of risk in their reporting for each EAC.

These monthly probabilistic estimates provide managers and decision makers with the level of certainty of achieving an EAC and the risk of overrunning it. This can be a valuable assessment mechanism for a PM to evaluate the estimate received from contractors on a monthly basis. If a PM receives an estimate from a contractor that is outside of the set upper bound, this discord could be investigated further and remedied if necessary. Also, by comparing the current month's EAC CDF to the previous month's EAC CDF, a PM could assess the changes in risk level between the two months. A notional example using Figure 16 will help illustrate this use.

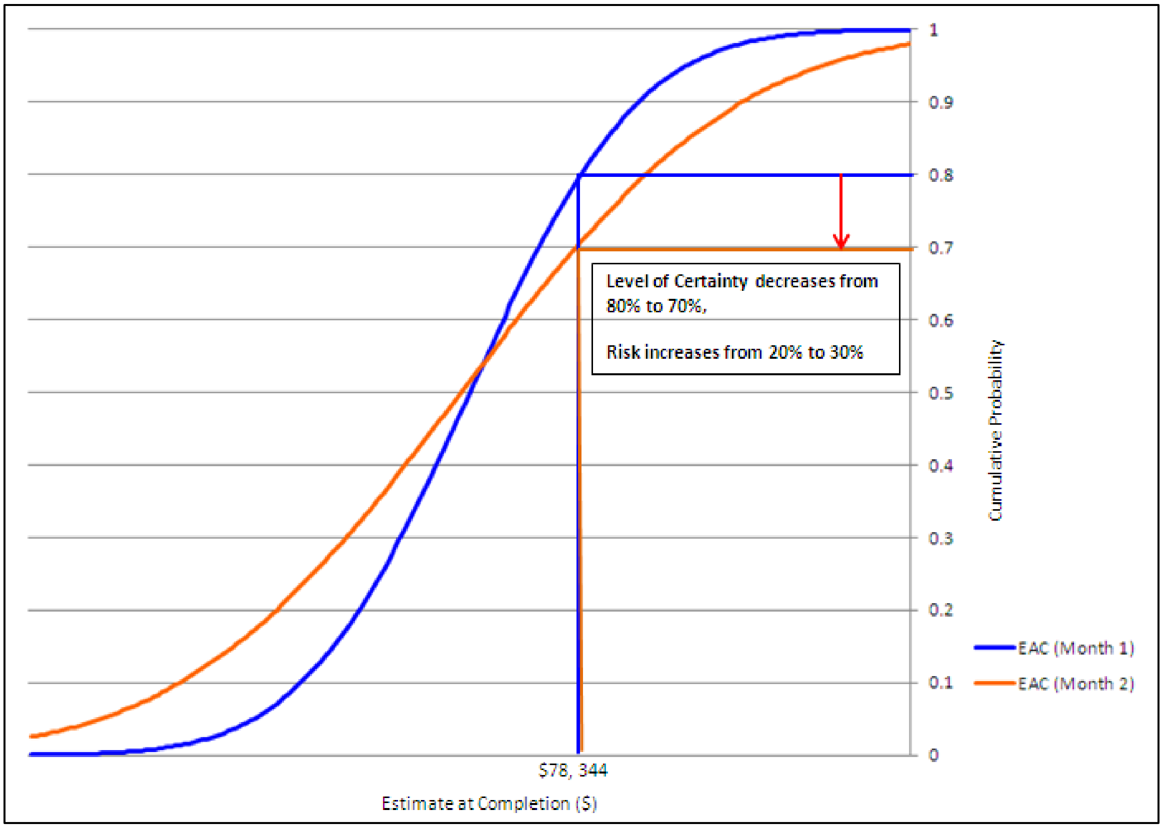

Figure 16. Comparison of monthly probabilistic EACs.

In Figure 16, the EAC at 80% level of certainty for month 1 is $78,344. In month 2, the PM would run the model with month 2's performance indices and determine the corresponding level of certainty associated with EAC = $78,344. In this case, the PM would learn that the level of certainty decreases from 80% to 70%, and thus the corresponding risk in the estimate has increased from 20% to 30%. This information provides the PM with a general sense of the cost risk trends from month to month based on the EV performance indices. Instead of best and worst case point estimate EACs alone, the PM now has a probabilistic model that provides EACs with associated levels of risk.

Vargas' probabilistic model makes use of the existing EV data and provides PMs with higher quality information. It allows the benefits of the cost analysis process to be continued throughout the life of the contract using EV data only. While the Vargas model may not achieve the same level of accuracy of the CEAC (because the Vargas model is limited to triangular distributions which may not correctly reflect the uncertainty associated with individual WBS cost elements), it offers a low cost and quick way of establishing a probabilistic estimate and improves the quality of information available to program managers. Moreover, all programs using EVM could subscribe to this method easily without unnecessarily burdening or restricting program management teams because the information is already available. The only additional investment would be the time to create the model.

Using cost uncertainty analysis methods, cost estimators can state with confidence the likelihood of exceeding their initial estimate (Garvey 2000). In a similar way, both the CEAC and Vargas methods provide the same risk information as the initial cost estimate. This thesis suggests using more of the probabilistic EAC output from these models in formal reporting and as an assessment mechanism. This will resolve ambiguity associated with deterministic estimates, help program managers and decision makers assess the uncertainty in the estimate, and help quantify the cost risk associated with the contract. Additionally, by standardizing the levels of certainty associated with best case, worst case, and most likely in EAC formal reporting, the DoD will improve its ability to make comparisons within and across programs. Congress and DoD officials would be better equipped to assess the risks of exceeding the estimate as well—an additional piece of the cost picture unavailable to them in the currently required reports. Overall, these small changes improve the quality of the information to the program manager and decision makers and require little additional investment to implement.

3. Improved Probit Model Using Probabilistic EAC to Forecast Likelihood of Program Cancellation

The probit models developed in Chapter IV were predicated on two EV variables to forecast the likelihood of cancellation: program manager’s and contractor’s EAC growth at 25%. The preceding sections of this chapter have shown how to improve the quality of the EAC from a deterministic estimate to a probabilistic one with associated levels of certainty. These probabilistic estimates at a chosen level of certainty can easily be used to measure cost growth in the same way the deterministic estimates were used. This thesis now suggests giving the program manager and decision maker an additional tool using these probabilistic EACs in the probit model to forecast likelihood of program cancellation. Figure 17 shows how the probabilistic EACs transformed into cost growth from two completion points of the contract can be plotted on the probit model to forecast the likelihood of cancellation.

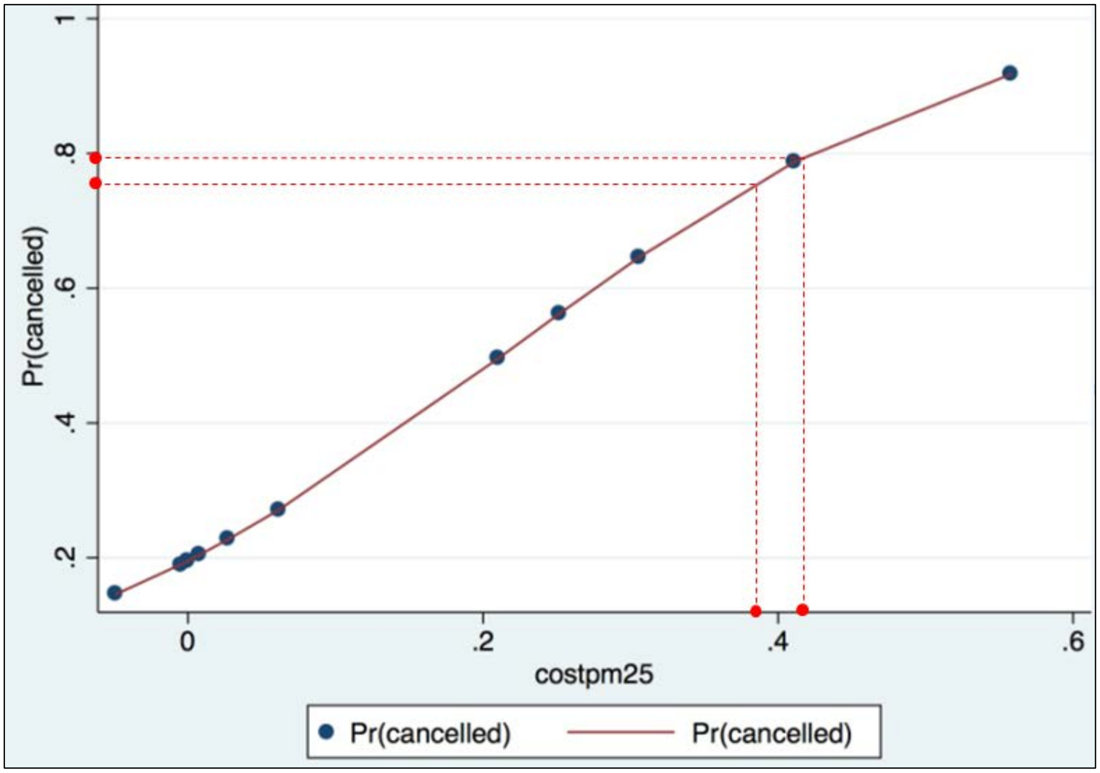

Figure 17. Probit model with probabilistic cost growth.

For example, using the values in Table 23, say a program manager is interested in the likelihood of cancellation and wants to be 80–90% certain of the forecast. The PM would take the associated EACs for these two levels of certainty and calculate the cost growth using the EACs from the last reporting period of the same level of certainty (as was demonstrated in Chapter III of this thesis). Hypothetically, the PM may calculate EAC growth of 38% and 42% respectively. Then, the PM would plot the EAC growths on the x-axis, go up to the probit S-curve and then over to the y-axis to learn that the program has between a 78–80% likelihood of being cancelled. This tool can be used quickly and cheaply as a forecasting tool by senior DoD officials and program managers to better understand the likelihood of cancellation due to early cost growth in a program.

D. SUMMARY

This thesis suggests that higher early cost growth in the form of contractor's and program manager's estimates at completion indicates a greater likelihood of program cancellation. The importance of the EAC led to a study on how to improve the use of probabilistic EACs in current reporting so that they are more valuable to program managers and decision makers. This study also recommends using the probabilistic EAC with this thesis' probit model to forecast the likelihood of cancellation.

VII. CONCLUSION

A. SUMMARY

The primary objectives of this thesis were two-fold: (1) to investigate whether there are differences in the key earned value metrics of cancelled and troubled non-cancelled programs, including the effect that resetting variances may have on program survival, and (2) to develop a model that captures the probability of cancellation based on earned value information. The analysis indicates that of the troubled programs sampled, cancelled programs do not have a lower or more unfavorable cost variance percentage or schedule variance percentage than troubled non-cancelled programs. Furthermore, the results suggest that cancelled programs do not have greater disparity in their contractor and program manager's estimates at completion, nor do they have greater cost and schedule variance reset frequencies. The Mann Whitney testing does suggest, however, that the cancelled programs have higher total cost growth over the life of the contract based on contractor EAC and program manager EAC. Probit modeling confirmed that cost growth at the 25% completion point for both the contractor and program manager EAC was greater for cancelled programs. Taken together, these results suggest that there is reasonably strong evidence that early cost growth is an indicator of program cancellation. A probit model captured the likelihood of program cancellation based on the EAC at the 25% completion point. These findings confirmed the importance of controlling cost growth early in a program.

The DoD has acknowledged that EVM is one of industry's and DoD's most powerful program management tools and it has recommitted itself to improving EVM implementation in MDAPs (Under Secretary of Defense for Acquisition, Technology and Logistics 2011). In fact, the DoD has recently demonstrated in April 2013 its seriousness and commitment to EVM by withholding 5% ($130M) of progress payments to a defense contractor over flaws in their earned value management system (Capaccio 2013). This study's findings and recommendations can help achieve EVM's overarching goal of transforming data into useful information for decision makers and support the DoD's initiative to improve the effectiveness of EVM. The recommendations of this thesis are,

(1) focused on the strongest finding of this report that cost growth in terms of EAC is an indicator of program cancellation, and (2) predicated upon the notion that probabilistic estimates that contain cost overrun risk information offer higher quality information and are more useful to decision makers (U.S. Government Accountability Office 2009). First, there should be more uncertainty information available in the annual Comprehensive Estimate at Completion in the SAR reporting of MDAPs. The benefit of this additional information is that Congress would now have a range of probabilistically determined EACs with associated levels of overrun risk to make cost growth assessments within programs and comparisons between programs. Second, consider using the Vargas probabilistic approach to calculating estimates throughout the life of the contract. This approach offers systems engineers and program managers additional insight into cost trends and overrun risk by providing monthly estimate and risk of overrun information with little additional effort required to construct the model. Third, consider using probit modeling to better understand the impacts of early cost overruns on the likelihood of program cancellation. The DoD has established PARCA to ensure that EV data is accurate, reliable, and timely, and that EVM is implemented in a disciplined manner (Under Secretary of Defense for Acquisition, Technology and Logistics 2011). The above recommendations help PARCA in their endeavor to achieve these standards. All of these recommendations improve the quality and realism of EV information presented to systems engineers, program managers, DoD officials and Congress and afford better management, budgeting, and cancellation decision making from a cost growth perspective.

B. AREAS FOR FURTHER RESEARCH

This research used a relatively small data set in its analysis. Future studies could increase the number of cancelled and non-cancelled programs and explore additional EVM performance metrics resulting in potentially richer analysis, new and stronger findings and higher fidelity probit models capable of better forecasting the likelihood of program cancellation. In addition to opening the data aperture, some of the variables explored in this research could be approached from different angles. This research found that cancelled programs did not have more cost or schedule variance resets than their non-cancelled counterparts. Further analysis could examine the size and timing of these resets and their relationship to program cancellation using survival analysis. This study also found that the difference between contractor EAC and program manager EAC was not an indicator of program cancellation. The relatively low p-value associated with this finding, however, suggests an area for further investigation of this potential discriminating variable. Lastly, additional study comparing the probabilistic EAC results obtained using the Vargas method with actual contract data would assess the accuracy of the model and determine its value to program management.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. MANN WHITNEY TEST RESULTS

This appendix contains the statistical results of the Mann Whitney tests that were conducted.

A Mann Whitney test (alpha = 0.10) tested the null hypothesis, H_0: $\mathrm{m_C} = \mathrm{m_{NC}}$ that the median CV% for cancelled and non-cancelled programs are equal. The alternative hypothesis is that the median CV% for cancelled programs is less (i.e., more negative and thus more unfavorable) than the median cost variance percentage of non-cancelled programs, H_A: $\mathrm{m_C} < \mathrm{m_{NC}}$. Figure 18 contains the Mann Whitney test results.

Mann-Whitney Test and CI: Cancelled CV%, Non-Canc CV%

```
                N  Median
Cancelled CV%   8   -2.59
Non-Canc CV%    8   -5.58


Point estimate for ETA1-ETA2 is 2.83
91.7 Percent CI for ETA1-ETA2 is (-0.65,6.20)
W = 84.0
Test of ETA1 = ETA2 vs ETA1 < ETA2


Cannot reject since W is > 68.0
```

Figure 18. Mann Whitney test for cost variance.

The null hypothesis for this test is that the median CV% for cancelled and non-cancelled programs are equal, H_0: $\mathrm{m_C} = \mathrm{m_{NC}}$. The alternative hypothesis is that the median CV% for the non-cancelled programs is less (more negative) than the median CV% for the cancelled programs, H_A: $\mathrm{m_{NC}} < \mathrm{m_C}$. Figure 19 contains the Mann Whitney test results.

Mann-Whitney Test and CI: Non-Canc CV%, Cancelled CV%

```
               N  Median
Non-Canc CV%   8   -5.58
Cancelled CV%  8   -2.59


Point estimate for ETA1-ETA2 is -2.83
91.7 Percent CI for ETA1-ETA2 is (-6.20,0.65)
W = 52.0
Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.0518
The test is significant at 0.0517 (adjusted for ties)
```

Figure 19. Second Mann Whitney test for cost variance.

The null hypothesis for this test is that the median CV% at the 25% contract completion point for cancelled and non-cancelled programs is equal, $H_0\colon \mathrm{m_C} = \mathrm{m_{NC}}$. The alternative hypothesis is that the median CV% at the 25% completion point for cancelled programs is less (i.e., more negative) than the median CV% at the 25% completion point for non-cancelled programs, $H_A\colon \mathrm{m_C} < \mathrm{m_{NC}}$. Figure 20 contains the test results.

Mann-Whitney Test and CI: Canc. CV @ 25%, Non-Canc. CV% at 25%

```
                     N  Median
Canc. CV @ 25%       7   -6.18
Non-Canc. CV% at 25% 8   -4.42


Point estimate for ETA1-ETA2 is -4.65
90.7 Percent CI for ETA1-ETA2 is (-13.72,2.33)
W = 47.0
Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.1626
```

Figure 20. Mann Whitney test for CV% at 25% contract completion point.

The null hypothesis for this test is that the median CV% at the 50% contract completion point for cancelled and non-cancelled programs is equal, $H_0\colon \mathrm{m_C} = \mathrm{m_{NC}}$. The alternative hypothesis is that the median CV% at the 50% completion point for cancelled programs is less (i.e., more negative) than the median CV% at the 50% completion point for non-cancelled programs, $H_A\colon \mathrm{m_C} < \mathrm{m_{NC}}$. Figure 21 contains the results.

Mann-Whitney Test and CI: Canc. CV @ 50%, Non-Canc. CV% at 50%

```
                       N  Median
Canc. CV @ 50%         7   -4.96
Non-Canc. CV% at 50%   8   -8.96


Point estimate for ETA1-ETA2 is 2.58
90.7 Percent CI for ETA1-ETA2 is (-3.45,6.65)
W = 64.0
Test of ETA1 = ETA2 vs ETA1 < ETA2


Cannot reject since W is > 56.0
```

Figure 21. Mann Whitney test for CV% at 50% contract completion point.

The null hypothesis for this test is that the median CV% at the 50% contract completion point for cancelled and non-cancelled programs is equal, $H_0\colon \mathrm{m_C} = \mathrm{m_{NC}}$. The alternative hypothesis is that the median CV% at the 50% completion point for non-cancelled programs is less (i.e., more negative) than the median CV% at the 50% completion point for cancelled programs, $H_A\colon \mathrm{m_{NC}} < \mathrm{m_C}$. Figure 22 contains the results of this test.

Mann-Whitney Test and CI: Non-Canc. CV% at 50%, Canc. CV @ 50%

```
                       N  Median
Non-Canc. CV% at 50%   8   -8.96
Canc. CV @ 50%         7   -4.96


Point estimate for ETA1-ETA2 is -2.58
90.7 Percent CI for ETA1-ETA2 is (-6.65,3.45)
W = 56.0
Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.1927
```

Figure 22. Second Mann Whitney test for CV % at 50% contract completion point.

A Mann Whitney test (alpha = 0.10) tested the null hypothesis, $H_0\colon \mathrm{m_C} = \mathrm{m_{NC}}$, that the median SV% for cancelled and non-cancelled programs are equal. The alternative hypothesis is that the median SV% for cancelled programs is less than (i.e., more negative and thus more unfavorable) than the median SV% of non-cancelled programs, $H_A: \mathrm{m_C} < \mathrm{m_{NC}}$. Figure 23 contains the Mann Whitney test results.

Mann-Whitney Test and CI: Cancelled SV %, Non-Canc SV%

```
                 N  Median
Cancelled SV %   8  -2.060
Non-Canc SV%     8  -3.065


Point estimate for ETA1-ETA2 is 0.855
91.7 Percent CI for ETA1-ETA2 is (-1.396,3.886)
W = 78.0
Test of ETA1 = ETA2 vs ETA1 < ETA2


Cannot reject since W is > 68.0
```

Figure 23. Mann Whitney test for schedule variance.

A Mann Whitney test (alpha = 0.10) tested the null hypothesis, $H_0: \mathrm{m_C} = \mathrm{m_{NC}}$, that the median SV% for cancelled and non-cancelled programs are equal. The alternative hypothesis is that the median SV% for non-cancelled programs is less than (i.e., more negative and thus more unfavorable) than the median SV% of cancelled programs, $H_A: \mathrm{m_C} < \mathrm{m_{NC}}$. Figure 24 contains the Mann Whitney test results.

Mann-Whitney Test and CI: Cancelled SV %, Non-Canc SV%

```
                 N  Median
Cancelled SV %   8  -2.060
Non-Canc SV%     8  -3.065


Point estimate for ETA1-ETA2 is 0.855
91.7 Percent CI for ETA1-ETA2 is (-1.396,3.886)
W = 78.0
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.1592
```

Figure 24. Second Mann Whitney test for schedule variance.

A Mann Whitney test (alpha = 0.10) tested the null hypothesis, $H_0\!: \mathrm{m_C} = \mathrm{m_{NC}}$, that the median SV% at the 25% completion point for cancelled and non-cancelled programs are equal. The alternative hypothesis is that the median SV% for cancelled programs at the 25% completion point is less than (i.e., more negative and thus more unfavorable) than the median SV% of non-cancelled programs, $H_A\!: \mathrm{m_C} < \mathrm{m_{NC}}$. The results of this test are found in Figure 25.

Mann-Whitney Test and CI: Canc. SV @ 25%, Non-Canc. SV @ 25%

```
                    N  Median
Canc. SV @ 25%      6   -6.18
Non-Canc. SV @ 25%  8   -4.89


Point estimate for ETA1-ETA2 is -1.20
91.9 Percent CI for ETA1-ETA2 is (-6.82,5.62)
W = 42.0
Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.3734
```

Figure 25. Mann Whitney test for SV% at 25% contract completion point.

The average of the SV% within +/- 5% of 50% completion point was computed and used for the schedule variance percentage of each program. A Mann Whitney test (alpha = 0.10) tested the null hypothesis, $H_0\!: \mathrm{m_C} = \mathrm{m_{NC}}$, that the median SV% at the 50% completion point for cancelled and non-cancelled programs are equal. The alternative hypothesis is that the median SV% for cancelled programs at the 50% completion point is less than (i.e., more negative and thus more unfavorable) than the median SV% of non-cancelled programs, $H_A\!: \mathrm{m_C} < \mathrm{m_{NC}}$. The results of this test are found in Figure 26.

Mann-Whitney Test and CI: Canc. SV @ 50%, Non-Canc. SV @ 50%

	N	Median
Canc. SV @ 50%	5	-4.44
Non-Canc. SV @ 50%	8	-5.41

Point estimate for ETA1-ETA2 is -0.10
90.8 Percent CI for ETA1-ETA2 is (-10.84,4.57)
W = 35.0
Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.5000

Figure 26. Mann Whitney test for SV% at 50% contract completion point.

A Mann Whitney test (alpha = 0.10) tested the null hypothesis, $H_0\text{: m}_{\text{C}} = \text{m}_{\text{NC}}$, that the median total growth of the program manager's EAC for cancelled and non-cancelled programs is equal. The alternative hypothesis is that the median total growth of the program manager's EAC for cancelled programs is greater than those of non-cancelled programs, $H_A\text{: m}_{\text{C}} > \text{m}_{\text{NC}}$. Figure 27 contains the Mann Whitney test results.

Mann-Whitney Test and CI: Cancelled PMEAC Cum Growth, Non-Canc PMEAC Cum Growth

	N	Median
Cancelled PMEAC Cum Growth	8	0.899
Non-Canc PMEAC Cum Growth	8	0.458

Point estimate for ETA1-ETA2 is 0.351
91.7 Percent CI for ETA1-ETA2 is (-0.160,0.927)
W = 80.0
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.1136

Figure 27. Mann Whitney test for median total growth of program manager's EAC.

The null hypothesis for the first test is that the median cumulative cost growth at the 25% contract completion point for cancelled and non-cancelled programs is equal, $H_0\text{: m}_{\text{C}} = \text{m}_{\text{NC}}$. The alternative hypothesis is that the median cumulative cost growth at the 25% completion point is greater for cancelled programs than for non-cancelled programs, $H_A\text{: m}_{\text{C}} > \text{m}_{\text{NC}}$. Figure 28 contains the test results.

Mann-Whitney Test and CI: Canc. PMEAC Growth @ 25%, Non-Canc. PMEAC Growth @ 25

```
                             N  Median
Canc. PMEAC Growth @ 25%     6   0.331
Non-Canc. PMEAC Growth @ 25% 8   0.004


Point estimate for ETA1-ETA2 is 0.251
91.9 Percent CI for ETA1-ETA2 is (0.034,0.558)
W = 59.0
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.0407
The test is significant at 0.0405 (adjusted for ties)
```

Figure 28. Mann Whitney test for cumulative cost growth (program manager's estimate at completion) at 25% contract completion point.

A Mann Whitney test with alpha = 0.10 was utilized to determine whether a difference in the samples' cumulative cost growth at contract completion percentage of 50% existed between the sampled cancelled and non-cancelled programs. The null hypothesis for this test is that the median cumulative cost growth at the 50% contract completion point for cancelled and non-cancelled programs is equal, $H_0: m_C = m_{NC}$. The alternative hypothesis is that the median cumulative cost growth at the 50% completion point is greater for cancelled programs than for non-cancelled programs, $H_A: m_C > m_{NC}$. Figure 29 contains the test results.

Mann-Whitney Test and CI: Canc. PMEAC Growth @ 50%, Non-Canc. PMEAC Growth @ 50

```
                             N  Median
Canc. PMEAC Growth @ 50%     5   0.472
Non-Canc. PMEAC Growth @ 50% 8   0.164


Point estimate for ETA1-ETA2 is 0.235
90.8 Percent CI for ETA1-ETA2 is (-0.068,0.882)
W = 45.0
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.0822
```

Figure 29. Mann Whitney test for cumulative cost growth (program manager's estimate at completion) at 50% contract completion point.

The null hypothesis can be rejected at the 10% level of significance, since the p-value = 0.0822 is less than 0.10.

A Mann Whitney test (alpha = 0.10) tested the null hypothesis, H_0: $m_C = m_{NC}$, that the median total growth of the contractor's EAC for cancelled and non-cancelled programs is equal. The alternative hypothesis is that the median total growth of the contractor's EAC for cancelled programs is greater than those of non-cancelled programs, H_A: $m_C > m_{NC}$. Figure 30 contains the Mann Whitney test results.

Mann-Whitney Test and CI: Cancelled Kor EA, Non-Canc Kor EAC

```
                             N  Median
Cancelled Kor EAC Cum Growth  8   0.917
Non-Canc Kor EAC Cum Growth   8   0.323


Point estimate for ETA1-ETA2 is 0.509
91.7 Percent CI for ETA1-ETA2 is (-0.027,0.982)
W = 83.0
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.0639
```

Figure 30. Mann Whitney test for median total growth of contractor's estimate at completion.

The null hypothesis can be rejected at the 10% level of significance, since the p-value = 0.0639 is less than alpha = 0.10.

A Mann Whitney test with alpha = 0.10 was used to determine whether a difference in the samples' cumulative cost growth at contract completion levels of 25% and 50% (again, there was not enough data available at the 75% completion point for sufficient analysis) existed between the sampled cancelled and non-cancelled programs. The null hypothesis for the first test is that the median cumulative cost growth at the 25% contract completion point for cancelled and non-cancelled programs is equal, $H_0 : m_C = m_{NC}$. The alternative hypothesis is that the median cumulative cost growth at the 25% completion point for cancelled programs is greater than the median cumulative cost growth at the 25% completion point for non-cancelled programs, $H_A : m_C > m_{NC}$. Figure 31 contains the Mann Whitney test results.

Mann-Whitney Test and CI: Canc. KorEAC Gro, Non-Canc. KorEAC

```
                                N  Median
Canc. KorEAC Growth @ 25%       6   0.327
Non-Canc. KorEAC Growth @ 25%   8   0.032


Point estimate for ETA1-ETA2 is 0.240
91.9 Percent CI for ETA1-ETA2 is (-0.005,0.559)
W = 58.0
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.0533
The test is significant at 0.0531 (adjusted for ties)
```

Figure 31. Mann Whitney test for cumulative cost growth (contractor's EAC) at 25% contract completion point.

The null hypothesis can be rejected at the 10% level of significance, since the p-value = 0.0533 is less than alpha = 0.10.

The null hypothesis for this test is that the median cumulative cost growth at the 50% contract completion point for cancelled and non-cancelled programs is equal, $H_0 : \mathrm{m_C} = \mathrm{m_{NC}}$. The alternative hypothesis is that the median cumulative cost growth at the 50% completion point for cancelled programs is greater than the median cumulative cost growth at the 50% completion point for non-cancelled programs, $H_A : \mu_{\mathrm{C}} > \mu_{\mathrm{NC}}$. Figure 32 contains the test results.

Mann-Whitney Test and CI: Canc. KorEAC Gro, Non-Canc. KorEAC

```
                                N  Median
Canc. KorEAC Growth @ 50%       5   0.528
Non-Canc. KorEAC Growth @ 50%   8   0.203


Point estimate for ETA1-ETA2 is 0.417
90.8 Percent CI for ETA1-ETA2 is (0.096,0.929)
W = 49.0
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.0241
```

Figure 32. Mann Whitney test for cumulative cost growth (contractor's EAC) at 50% contract completion point.

The null hypothesis can be rejected at the 10% level of significance, since the p-value = 0.0241 is less than alpha = 0.10.

A Mann Whitney test (alpha = 0.10) tested the null hypothesis, $H_0\colon \mathrm{m_C} = \mathrm{m_{NC}}$, that the median EAC difference (between contractor and program manager estimates) for cancelled and non-cancelled programs is equal. The alternative hypothesis is that the median EAC difference for cancelled programs is greater than the median EAC difference of non-cancelled programs, $H_A\colon \mathrm{m_C} > \mathrm{m_{NC}}$. Figure 33 contains the Mann Whitney test results.

Mann-Whitney Test and CI: Canc EAC Difference, Non-Canc EAC Difference

```
                          N  Median
Canc EAC Difference       8  0.0209
Non-Canc EAC Difference   8  0.0351


Point estimate for ETA1-ETA2 is -0.0088
91.7 Percent CI for ETA1-ETA2 is (-0.0445,0.0203)
W = 62.0
Test of ETA1 = ETA2 vs ETA1 > ETA2


Cannot reject since W is < 68.0
```

Figure 33. Mann Whitney test for estimate at completion difference.

A second test was conducted to see if non-cancelled EAC difference was worse than cancelled EAC difference and the results appear in Figure 34.

Mann-Whitney Test and CI: Canc EAC Difference, Non-Canc EAC Difference

```
                          N  Median
Canc EAC Difference       8  0.0209
Non-Canc EAC Difference   8  0.0351


Point estimate for ETA1-ETA2 is -0.0088
91.7 Percent CI for ETA1-ETA2 is (-0.0445,0.0203)
W = 62.0
Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.2818
The test is significant at 0.2816 (adjusted for ties)
```

Figure 34. Second Mann Whitney test for estimate at completion difference.

Since the p-value of 0.2816 greatly exceeds the alpha = 0.10, the null hypothesis can not be rejected in this case either.

To further investigate the effect of cost variance resets, a Mann Whitney test (alpha = 0.10) tested the null hypothesis, $H_0\text{: m}_{\text{C}} = \text{m}_{\text{NC}}$, that the median cost variance reset frequencies for cancelled and non-cancelled programs are equal. The alternative hypothesis is that the median cost variance reset frequencies for cancelled programs is greater than those of non-cancelled programs, $H_A : \text{m}_{\text{C}} > \text{m}_{\text{NC}}$. The purpose of this test is to investigate whether cancelled programs have a greater frequency of cost variance resets than non-cancelled programs. If this is the case, it may be a leading indicator to predict cancellation. The results contained in Figure 35 revealed a p-value of 0.2474 for cost variance resets and therefore not enough evidence exists in the data to reject the null hypothesis.

Mann-Whitney Test and CI: Canc. CV Reset Freq., Non-Canc. CV Reset Freq.

```
                          N  Median
Canc. CV Reset Freq.      8   1.000
Non-Canc. CV Reset Freq.  8   0.000


Point estimate for ETA1-ETA2 is 0.000
91.7 Percent CI for ETA1-ETA2 is (-0.000,1.000)
W = 75.0
Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.2474
The test is significant at 0.2260 (adjusted for ties)
```

Figure 35. Mann Whitney test for cost variance reset frequency.

A similar Mann Whitney test was conducted comparing the frequency of schedule variance resets. The test results are contained in Figure 36.

Mann-Whitney Test and CI: Canc. SV Reset Freq., Non-Canc. SV Reset Freq.

```
                           N  Median
Canc. SV Reset Freq.       8   0.000
Non-Canc. SV Reset Freq.   8   0.000


Point estimate for ETA1-ETA2 is -0.000
91.7 Percent CI for ETA1-ETA2 is (-1.000,1.000)
W = 66.5
Test of ETA1 = ETA2 vs ETA1 < ETA2 is significant at 0.4582
The test is significant at 0.4516 (adjusted for ties)
```

Figure 36.   Mann Whitney test for schedule variance frequency.

The resulting p-value of 0.4582 indicates that there is little evidence to reject the null hypothesis.

APPENDIX B. PROBIT RESULTS

Appendix B contains the results of the individual probit regression tests.

VARIABLES	(1) cancelled	(2) cancelled	(3) cancelled	(4) cancelled	(5) cancelled	(6) cancelled	(7) cancelled	(8) cancelled	(9) cancelled	(10) cancelled	(11) cancelled
CV %	0.186										
	(0.151)										
CV % at 25% completion		-0.0666									
		(0.186)									
CV % at 50% completion			0.0159								
			(0.791)								
CV reset frequency				0.326							
				(0.473)							
SV%					-0.0186						
					(0.796)						
SV % at 25% completion						-0.0115					
						(0.764)					
SV % at 50% completion							-0.0656				
							(0.373)				
SV reset frequency								-0.127			
								(0.751)			
Cost Growth (PMEAC)									0.624		
									(0.256)		
PMEAC at 25% complete										4.029*	
										(0.0900)	
PMEAC at 50% complete											1.446
											(0.251)
Constant	0.791	-0.389	0.0416	-0.204	-0.0783	-0.284	-0.670	0.0711	-0.442	-0.858*	-0.951
	(0.185)	(0.352)	(0.942)	(0.632)	(0.857)	(0.557)	(0.216)	(0.854)	(0.346)	(0.0842)	(0.123)
Observations	16	15	15	16	16	14	13	16	16	14	13
Pseudo R-squared	0.166	0.102	0.00335	0.0234	0.00310	0.00475	0.0510	0.00456	0.0918	0.280	0.181

pval in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 24. Probit results with Crusader (part 1).

VARIABLES	(1) cancelled	(2) cancelled	(3) cancelled	(4) cancelled	(5) cancelled	(6) cancelled	(7) cancelled
Cost Growth (Contractor EAC)	1.449						
	(0.102)						
Contractor EAC @ 25%		4.454*					
		(0.0954)					
Contractor EAC @ 50%			1.757				
			(0.163)				
Diff. b/w Contr. EAC & PMEAC				2.290			
				(0.577)			
Diff. b/w Contr. EAC & PMEAC (0-25%)					2.477		
					(0.706)		
Diff. b/w Contr. EAC & PMEAC (26-50%)						-6.431	
						(0.440)	
Diff. b/w Contr. EAC & PMEAC (51-75%)							2.225
							(0.633)
Constant	-0.916	-0.859*	-1.168*	-0.109	-0.277	0.115	-0.303
	(0.131)	(0.0801)	(0.0875)	(0.765)	(0.513)	(0.821)	(0.470)
Observations	16	14	13	16	14	14	14
Pseudo R-squared	0.202	0.293	0.236	0.0159	0.00761	0.0317	0.0128

pval in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 25. Probit results with Crusader (part 2).

VARIABLES	(1) cancelled	(2) cancelled	(3) cancelled	(4) cancelled	(5) cancelled	(6) cancelled	(7) cancelled	(8) cancelled	(9) cancelled	(10) cancelled	(11) cancelled
CV %	0.168										
	(0.193)										
CV % at 25% completion		-0.0654									
		(0.195)									
CV % at 50% completion			0.0103								
			(0.864)								
CV reset frequency				0.266							
				(0.561)							
SV%					-0.0341						
					(0.651)						
SV % at 25% completion						-0.00941					
						(0.805)					
SV % at 50% completion							-0.0391				
							(0.586)				
SV reset frequency								-0.193			
								(0.637)			
Cost Growth (PMEAC)									0.544		
									(0.428)		
PMEAC at 25% complete										4.029*	
										(0.0900)	
PMEAC at 50% complete											1.411
											(0.309)
Constant	0.654	-0.484	-0.0980	-0.245	-0.233	-0.378	-0.641	0.0175	-0.401	-0.858*	-0.939
	(0.283)	(0.263)	(0.867)	(0.568)	(0.611)	(0.443)	(0.229)	(0.964)	(0.435)	(0.0842)	(0.146)
Observations	15	14	14	15	15	13	12	15	15	13	12
Pseudo R-squared	0.150	0.106	0.00152	0.0164	0.0105	0.00353	0.0205	0.0109	0.0307	0.205	0.0718

pval in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 26. Probit results without Crusader (part 1).

VARIABLES	(1) cancelled	(2) cancelled	(3) cancelled	(4) cancelled	(5) cancelled	(6) cancelled	(7) cancelled
Cost Growth (Contractor EAC)	1.449						
	(0.102)						
Contractor EAC @ 25%		4.454*					
		(0.0954)					
Contractor EAC @ 50%			1.757				
			(0.163)				
Diff. b/w Contr. EAC & PMEAC				2.776			
				(0.511)			
Diff. b/w Contr. EAC & PMEAC (0-25%)					2.477		
					(0.706)		
Diff. b/w Contr. EAC & PMEAC (26-50%)						-4.379	
						(0.608)	
Diff. b/w Contr. EAC & PMEAC (51-75%)							3.275
							(0.504)
Constant	-0.916	-0.859*	-1.168*	-0.221	-0.277	-0.0808	-0.486
	(0.131)	(0.0801)	(0.0875)	(0.559)	(0.513)	(0.882)	(0.278)
Observations	15	13	12	15	14	13	13
Pseudo R-squared	0.146	0.219	0.134	0.0245	0.00761	0.0153	0.0289

pval in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 27. Probit results without Crusader (part 2).

LIST OF REFERENCES

Abba, Wayne. 1997. "Earned Value Management — Reconciling Government and Commercial Practices." *Program Manager* 26: 58–63.

Adams, Gordon. 2013. "Howling into the wind." *Foreign Policy*, March 4. http://adams.foreignpolicy.com/posts/2013/03/04/howling_into_the_wind.

Angelis, Diana I. 2012. "Cost Estimating Course Notes." SE3011 Cost Estimation, Spring Quarter, Monterey, CA.

Assistant Secretary of Defense for Acquisition, Office of the. 2012. "Integrated Program Management Report (IPMR) Data Item Description (DID)." Memorandum, June 19. Washington, DC.

Augustine, Norman R. 1997. *Augustine's Laws*. 6th ed. Reston: American Institute of Aeronautics and Astronautics.

Beach, Chester P., Jr. 1990. *A-12 Administrative Inquiry*. Washington, DC: Department of the Navy.

Berteau, David, Guy Ben-Ari, Joachim Hofbauer, Gregory Sanders, Jesse Ellman, and David Morrow. 2011. *Cost and Time Overruns for Major Defense Acquisition Programs*. Wasington, DC: Center for Strategic and International Studies.

Capaccio, Tony. 2013. "Pentagon Withholding From Lockheed Grows Over Management." *Bloomberg.com*, April 22. http://www.bloomberg.com/news /2013–04–22/pentagon-withholding-from-lockheed-grows-over-management.html.

Charette, Robert N. 2008. "What's Wrong with Weapons Acquisitions?" *IEEE Spectrum*, November 1. http://spectrum.ieee.org/aerospace/military/whats-wrong-with-weapons-acquisitions.

Chen, Yongmin, and Ron Smith. 2001. "Equilibrium Cost Overruns." *Annals of Economics and Finance* 2: 401–414.

Christensen, David D. 1994. "Cost Overrun Optimism: Fact or Fiction?" *Acquisition Review Quarterly* (Winter): 25–38.

———. 1999. "Using the Earned Value Cost Management Report to Evaluate the Contractor's Estimate at Completion." *Acquisition Review Quarterly* (Summer): 435.

Christensen, David S. 1993. "The Estimate at Completion Problem: A Review of Three Studies." *Project Management Journal* 24 (March): 37–42.

Defense Acquisition University. 2009. “Estimate at Completion.” https://dap.dau.mil/acquipedia/Pages/ArticleDetails.aspx?aid=684744f8–0de4–4d17-a5f0–84533ad5841a.

———. 2012a. “DAU Gold Card – July 2012.” https://acc.dau.mil/CommunityBrowser.aspx?id=19577.

———. 2012b. “Defense Acquisition Guidebook [DAG].” https://dag.dau.mil/Pages/Default.aspx.

———. 2012c. “EVM Statutes, Regulatory Policy, and DoD Implementation Guidance.” https://learn.dau.mil/.

Dubos, Gregory F., Joseph H. Saleh, and Robert Braun. 2007. “Technology Readiness Level , Schedule Risk and Slippage in Spacecraft Design: Data Analysis and Modeling.” Paper presented at the annual meeting for the American Institute of Aeronautics and Astronautics Space, Long Beach, CA, September 18–20.

Eisner, Howard. 2002. *Essentials of Project and Systems Engineering Management*, 2nd ed. New York: John Wiley & Sons.

Fast, William. 2013. “Acquisition Strategy.” Presentation, Naval Postgraduate School, Monterey, CA.

Garvey, Paul R. 2000. *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*. New York: Marcel Dekker.

Gates, Robert M. 2009. “Defense Budget Recommendation Statement.” Speech, Arlington, VA, April 6. http://www.defense.gov/Speeches /Speech.aspx?SpeechID=1341.

Hagan, Gary. 2009. *Defense Acquisition University Glossary of Defense Acquisition Acronyms and Terms*, 13th ed. Ft. Belvoir: Defense Acquisition Univ. Press.

Harrison, Todd. 2011. *Analysis of the FY12 Defense Budget*. Washington, DC: Center for Strategic and Budgetary Assessments.

Hosmer, David W. and Stanley Lemeshow. 1989. *Applied Logistic Regression*. New York: John Wiley & Sons.

Kerzner, Harold R. 2009. *Project Management: A Systems Approach to Planning, Scheduling, and Controlling*, 10th ed. Hoboken: John Wiley & Sons.

Koop, Gary. 2012. “Applied Econometrics: Topic 2. Logit and Probit Models.” http://personal.strath.ac.uk/gary.koop/ec408/EC408_Topic_2_slides.pdf.

Langford, Gary O., and Raymond Franck. 2009. *Gap Analysis: Application to Earned Value Analysis*, August 19. Technical Report (NPS-GSBPP-09–006). Monterey: Naval Postgraduate School.

Makielski, Marty. 2009. *Examining the Comprehensive Estimate-at-Completion*. Washington, DC: Defense Contracting Management Agency.

Management and Budget, Office of. 2006. *Capital Programming Guide* (June) (OMB Circular No. A-11). Washington, DC.

McKsymick, Eric M. 1995. *An Analysis of Cost Overruns in the Development of the Navy's A-12 Avenger Aircraft*. DTIC (ADA306100). Monterey: Naval Postgraduate School.

Methods Consultants of Ann Arbor. 2013. "What is the Difference between Probit and Logit Models?" http://www.methodsconsultants.com/tutorials/logprob.html.

Morrison, David C. 1991. "Deep-sixing the A-12." *Government Executive* 23 (March): 30–35.

Mun, Jonathan. 2012. *Risk Simulator User Manual*. Real Optics Valuation, Dublin, CA.

O'Halloran, Sharyn. 2013. "Lecture 9: Logit/Probit." http://www.columbia.edu/~so33/SusDev/Lecture_9.pdf.

Performance Assessments and Root Cause Analyses. 2013. "Earned Value Management." http://www.acq.osd.mil/evm/.

Sipple, Vince, Edward White, and Michael Greiner. 2004. *Surveying Cost Growth*. Ft. Belvoir: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics.

TutorialsPoint. 2013. "Earned Value Management Tutorial." http://www.tutorialspoint.com/earn_value_management.

Tyson, Karen. W., Bruce R. Harmon, and Daniel M. Utech. 1994. *Understanding Cost and Schedule Growth in Acquisition Programs* (IDA Paper P-2967). Alexandria: Institute for Defense Analyses.

U.S. Congress. 2009. *Weapon Systems Acquisition Reform Act of 2009.* Pub. L. No. 111–23. 111th Cong., 1st sess.

U.S. Department of Defense. 2007. *Defense Acquisition System*, November 20. DoD Directive 5000.01. Washington, DC: Under Secretary of Defense for Acquisition, Technology and Logistics.

———. 2008. *Operation of the Defense Acquisition System*, December 8. DoD Instruction 5000.02. Washington, DC: Under Secretary of Defense for Acquisition, Technology and Logistics.

U.S. Department of Defense. Defense Contract Management Agency. 2006. *Earned Value Management Implementation Guide*. Alexandria: Defense Contract Management Agency.

———. 2012. *Earned Value Management System (EVMS) Program Analysis Pamphlet (PAP)*, July 10. DCMA-EA PAM 200.1.

U.S. Government Accountability Office. 2009. *GAO Cost Estimating and Assessment Guide*. GAO-09–3SP. http://www.gao.gov/new.items/d093sp.pdf.

———. 2013. *Defense Acquisitions: Assessment of Selected Weapons Programs.* GAO-13–294SP. http://www.gao.gov/assets/660/653379.pdf.

Under Secretary of Defense for Acquisition, Technology and Logistics. 2011. "Earned Value Management Systems Performance, Oversight and Governance." Memorandum, January 24. Washington, DC.

University of California, Los Angeles: Statistical Consulting Group. 2013a. "Introduction to SAS." http://www.ats.ucla.edu/stat/sas/notes2/.

———. 2013b. "Stata Data Analysis Examples: Probit Regression." http://www.ats.ucla.edu/stat/stata/dae/probit.htm.

Vargas, Ricardo Viana. 2009. "Earned Value Probabilistic Forecasting Using Monte Carlo Simulation." http://www.ricardovargas.com.br.

Webb, Alan. 2003. *Using Earned Value: A Program Manager's Guide*. Burlington: Gower Publishing.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California